KTH Royal Institute of Technology

School of Electrical Engineering
Laboratory of Communication Networks

Characterizing WLAN Channel Occupancy for Cognitive Networking

Alexandre Vizcaino Luna

Advisor: Ioannis Glaropoulos
Examiner: Assoc. Prof. Viktoria Fodor

Stockholm
June 2012

ABSTRACT

The deployment of heterogeneous wireless networks in the same spectrum space introduces the need for dynamic spectrum access so as to increase the utilization of the available wireless resources. Dynamic spectrum access needs to be controlled in order avoid interference between the users of different systems. Different schemes can be used in order to avoid the mutual interference between the systems: orthogonality in space, frequency or time. In this thesis we address the problem with a solution based on time orthogonality, in which the coexisting wirless systems are a WLAN and WSN.

Due to the high power asymmetry it is necessary to implement a cognitive capability in the most affected system, i.e. the WSN, which will predict the behaviour of the WLAN spectrum usage and take advantage of the white spaces left for WSN interference-free communication. For this, it is necessary to model the traffic of the WLAN system. The applicability of two different semi-Markovian models has been studied in the scope of this thesis: one represents an ideal case in which the sensors have unlimited sensing capabilities and a second, more realistic, approach in which the sensor view is limited by hardware and resources. In this project we investigate whether and when the proposed models are suitable to be used in order to model, estimate and predict realistic WLAN channel usage; for that we consider a measurement-based, multi-layer WLAN traffic workload model.

Different experiments have been developed to test different traffic scenarios in which we apply our prediction model. The experiments show that the WLAN usage estimation process is robust, i.e. insensitive to irregularities introduced by the packet level randomization and the underlying protocols in the WLAN. An almost perfect fitting is achieved in a wide range of cases between the distributions to model the active and idle periods and the empirically derived channel usage functions. In addition, we study different usage load regions in which we apply our model and the results show that it can be applied with high success in a region of 10 to 30 % of load. On the other hand, the realistic model, based on partial observation of WLAN trafic, shows higher variations between different traffic conditions, increasing the performance of the estimation process in cases of higher WLAN load.

ACKNOWLEDGEMENTS

I would like to thank to all the people that made this project possible. First of all, I would like to deeply thank Ioannis Glaropoulos for his invaluable help and advice through all the project and the daily meetings discussing all the experiments and results, that allow me to learn a lot about the topic under study. My thanks also to Prof. Viktoria Fodor for her guidelines and suggestions for the experiments and for the detailed review of my thesis. Also my thanks to Marcello Laganà for his help with the software and the multiple suggestions to improve it. Most importantly, to my family for their support during the last year and for giving me the opportunity to study abroad.

CONTENTS

1 INTRODUCTION

The proliferation of new wireless communication systems had lead to a over-crowded unlicensed band, limiting the performance of the different systems due to the high interference environment. However, as some studies have shown, the spectrum is still lightly used [3]-[5] at most times and locations. This, in addition with the increase in the deployment of heterogeneous networks, leads to a necessity of new schemes to dynamically reuse the spectrum and increase the utilization of the resources available by introducing secondary systems. This increase in the usage of the spectrum will need to be controlled in order not to cause interference in the licensee (primary user) of the network. Mutual interference can be avoided in the design of the secondary system by using orthogonality in space, frequency or time [5]. In this document we will approach a solution based on orthogonality in time, in which the secondary system will take profit of the white spaces left from the communication of the primary system users.

In this project we will focus on WLAN users as primary system users and Wireless Sensor Networks WSN as secondary systems. The interference applied by the primary system over the second one is considerable and the transmission of the secondary system can be damaged by this interference. On the other hand, it is not necessary to take into account the interference of the secondary system to the primary since the transmit power of the primary user is much higher than the secondary user. The primary system communication leaves white spaces between active periods. In order to take profit from these white spaces in the spectrum left from the communication of the first system users and avoid the interference, it is necessary to predict when there will be a white space that can be used for the secondary system transmission and how long it will last. The interference of the primary system will have a high effect over the secondary system transmissions, which will cause a high number of collisions and therefore, force to retransmit the packets in order to complete the transmission. This last will affect the energy efficiency of the system, which is a critical issue due to the resource limitations in the WSN technology.

The solution under study in this project is the one introduced in [3] and then extended in [4] and [5]. These studies introduce as a solution a semi-Markovian model which presents a balance between complexity and accuracy, in order to model the WLAN traffic. However, their solution is applied in scenarios with low load of WLAN users. Our contribution is to study if and when the previous solution is applicable to different scenarios in order to model the traffic.

In order to model the traffic, we will use the model presented in [8] which develops a study of the session and flow levels of larger scenarios and traffic workload in a campus WLAN. In [10] the previous model is extended with the packet level by applying the semi-Markovian model introduced previously, conforming the final multi-layer traffic model that will be used in this project.

In this project, we will focus in the developing of a set of tests to study the fitness of the semi-Markovian model to model the WLAN traffic using a multi-layer traffic model as the one presented in the previous papers for different types of scenarios.

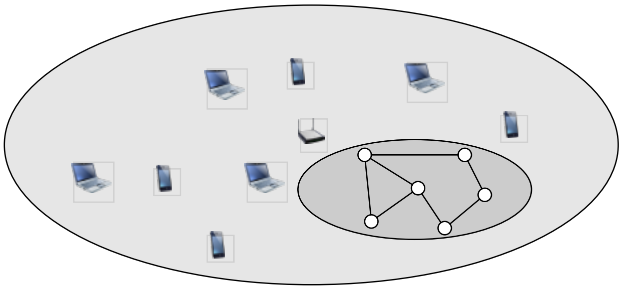


Figure 1.1: Networking Scenario

1.1 SCENARIO

As it has been introduced, the scenario under study in this project is a heterogeneous network in which WLAN and WSN technologies coexist in the shared spectrum, overlapping their communication bands. The model used during the tests consists in a single WLAN access point providing access to different WLAN users. Inside the WLAN coverage area a WSN is deployed (represented in Figure 1.1). The transmit power of the WLAN is much higher than the WSN system.

A cognitive access scheme (presented in [7]) is implemented in the sensor nodes in order to increase the efficiency to sense the channel and make the proper estimations of the traffic using the semi-Markovian model presented in [4].

The power consumption is a critical issue due to the hardware limitations in the WSN technology. For this, the sensors should be awake the shortest time possible while sensing the channel. During these awake periods, the sensor, using the cognitive capability implemented, should be able to sense the WLAN traffic. Then, offline, model the traffic with the semi-Markovian model and make the proper estimations in order to estimate the idle distribution. The prediction process presents a series of challenges: it should be fast and real-time in order to be able to predict the traffic behaviour before this changes. In addition, it should be as accurate as possible in order to model the traffic in the best way possible. The detailed process will be presented in following chapters. Since our aim is to study the sensing and estimation capabilities of the sensors, at the beginning we will not consider transmissions between the WSN nodes.

In addition, two different models for the sensor nodes are tested in function of the observable capabilities. An ideal case will be studied, in which the sensors are able to observe all the traffic of the network. On the other hand, another model will be also tested, in which the sensors only have the capacity of observe part of the network due to hardware limitations. For the last case, the 2-state semi-Markovian model has been extended in [10] to a 3-state semi-Markov model.

The multi-layer traffic model presented in [8] and extended in [10] is composed by the session, flow and packet levels as it is represented in Figure 1.2. In [8], it has been demonstrated that the session and flow levels can be approximated following some determined distributions with specific parameters. The configuration of these levels will be introduced later.

1.2 RELATED WORK ON WLAN SPECTRUM USAGE MODELING

Different studies have been developed around the issues that spectrum sharing in time-domain exploiting the white-space WS between transmissions in a WLAN 802.11 based network. In order to model the WLAN traffic, different proposals had been presented.

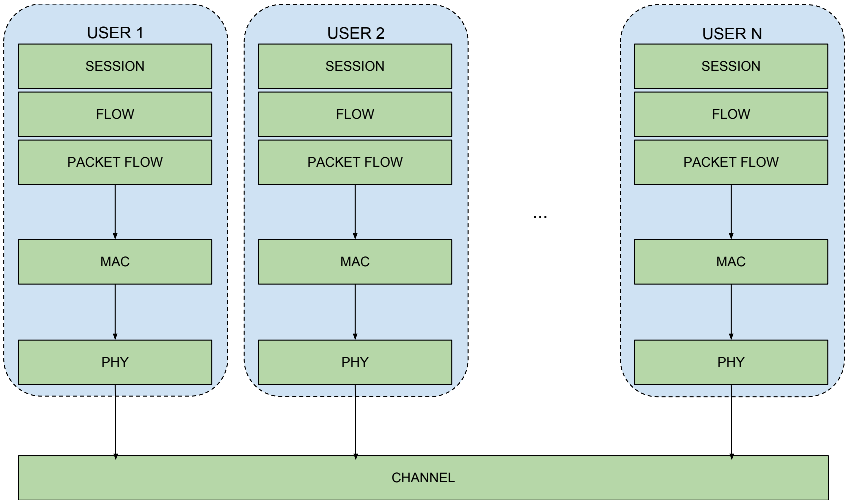


Figure 1.2: Multi-layer traffic model [8]

In [3], a first proposal of a continuous-time semi-Markovian model is presented in order to model the WLAN's channel behavior. This model presents a good trade-off between analytical tractability and accuracy. The aim is to implement a capability to sensors in order to predict the primary user's behavior by sensing the traffic generated by this primary user and later on generate a model using the proposed semi-Markovian model.

Due to the heavy-tail behavior of the idle periods, the continuous-time Markov process is not the proper model since the sojourn times in each state should be exponentially distributed. Because of this, they propose a semi-Markov model which allows an arbitrary specification of the sojourn time in each state. The sequence of states of transmission of Data - SIFS - ACK follows a deterministic behavior, which they decide to include them together as a single Active state. On the other hand, the white spaces between transmissions are defined as Idle state. Because the Idle state follows a heavy-tailed behavior, they consider a Generalized Pareto distribution to model this state. Even though, some problems appear with the Generalized Pareto distribution for high load scenarios.

Due to the problems presented in [3] because of the bad fitting for high load scenarios, in [4] the previous semi-Markovian model is extended by modifying the Idle distribution previously defined. In [3] is just considered that the idle periods are determined by the white-spaces of the transmissions. On the other hand, the contention window (CW) should be included also in the Idle state. For this, a mixture idle distribution to model the Idle periods of the channel is proposed. The mixture idle distribution consists in a combination of two distributions that model each the CW and WS. Since the CW follows an uniform behavior, an Uniform distribution is proposed as a solution. On the other hand, for the heavy-tail behavior of the WS, the Generalized Pareto is still a good solution to model this behavior but in this case, this distribution will be left-truncated due to the inclusion of the CW in the model.

In [8], the traffic of a Campus WLAN is studied in order to find the proper distributions to model it. Proposing a multi-layer traffic model where two different layers or levels of traffic are determined: session and flow. The packet level is not studied. In order to model each one of the levels, different distributions are proposed that fit the behaviour of each one. This multi-layer traffic model is extended with a new level for the packets in [10]. In addition, the 2-state semi-Markovian model is extended to a 3-state for the partially-observable model for the sensor nodes.

In [5], using as a base the model presented in [3] and [4], a Cognitive Medium Access (CMA) is introduced in order to be implemented over sensors to coexist with the WLAN. Two models are presented differentiated by the observable capability of the channel: full-observable and partially-observable, which will be a starting point in the two models presented in [7].

In [7] a Cognitive Access Scheme is presented for WSNs that coexist with WLANs, considering the blind and hidden WLAN terminals, in order to decrease the negative effect of the coexistence problem and normalize the energy cost considering the limited sensing capabilities of the sensor nodes. The mixture idle distribution proposed in [4] is used to model channel occupancy. The cognitive capabilities designed for the sensor nodes include: capability to differentiate if an idle period is due to CW or WS, decision capacity over the packet size and next hop distance, and predict whether there is sufficient WLAN idle time for the transmission of a WSN packet. The final results show that the proposed solution achieves a significant gain in performance compared to the traditional channel access solutions for typical WLAN load values.

In our project, the main goal will be extending the work developed in [10], designing a set of experiments to test the different implementations done and correct or extend when is needed. Also a set of experiments will be developed in order to test the complete semi-Markovian multi-layer traffic model for different traffic realizations in order to find in which situations the model is suitable.

1.3 THESIS CONTRIBUTION

Our main contribution in this project is to extend the work developed in [10]. In that project, the multi-layer traffic model presented in [8] is extended adding a packet level in addition with the extension of the two-state semi-Markovian model in [4] to a 3-state semi-Markovian model for the partially-observable model for the sensor nodes.

The implementation of the cognitive access scheme proposed in [7], and the multi-layer traffic model extended with the packet level is developed over NS2 simulation software. In [10] different experiments were developed to test the sensing and transmission capabilities of the WSN nodes, the estimation process for the Active and Idle distributions using Maximum-Likelihood Estimation and the implementation of the Kolmogorov-Smirnov test as a validation test.

In this project we will revise the work developed in [10] and correct/extend it for the different situations we can face. The main objective of this project is to design a full set of experiments in which the modelling is tested over different traffic realizations using the multi-layer traffic model in order to observe in which cases our model is suitable.

In addition, a Laplace Transform estimator will be implemented in order to estimate the parameters for the idle distribution in the Local View model.

The different experiments will be developed using the NS2 simulation software with the NSMiracle framework and the implementations developed in [10]. For this, different parts of the implementation developed in [10] should be tested in order to check the correct functionality of each of them before testing the final complete model. We need to focus in some key parts that need to be tested:

- Test of the mixture idle distribution proposed in [4].
- Test the estimation process of the parameters for the mixture idle distribution.
- Test the validation test chosen (Kolmogorov-Smirnov test) and choose a better one in case is needed.
- Test the final revised model over different traffic realizations using the multi-layer traffic model implemented on NS.

Chapter 2 will present the general model under study in this project, in addition with the Global View and Local View sensing models that will be used and their characteristics, presenting the scenario and estimation process for both models. Chapter 3 presents the multi-layer traffic model that will be used to generate the real traffic scenarios for our tests. Chapter 4 presents the main simulation characteristics, software to be used and main characteristics of the implementation developed in [9]. The results of the different experiments developed on our model will be presented in chapters 5 and 6. Finally, Chapter 7 summarizes the conclusions extracted from the different tests of our project.

Part I

MODELS

2 SPECTRUM OCCUPANCY MODEL

As it has been presented in Section 1.1, the model under study consists of a heterogeneous network composed by WLAN nodes and Wireless Sensor Nodes (WSN). Different issues are present in this type of scenario but the most significant one is the interference problem. Since the transmit power of the WSN is much lower than the transmit power used by the WLAN nodes, the WSN nodes are highly affected by the interference from the WLAN transmissions. For this, it is necessary to find the proper solution for this problem. For that, previous work proposed a Cognitive Layer in the WSN in order to give them the capability to sense the channel and predict when the network will be idle of transmissions. This prediction process will give the sensor the capability to model the network's spectrum occupancy behaviour and be able to transmit minimizing the present interference. In order to proceed with the prediction process, it is necessary to model the WLAN spectrum activity.

The WLAN spectrum activity is composed by different states that can be modelled. For this, a continuous time modelling it is necessary since WLAN does not have a slot structure. The model includes the transition between the channel's states and the duration of the time in each of the states [9]. This basic model is represented in Figure 2.1.

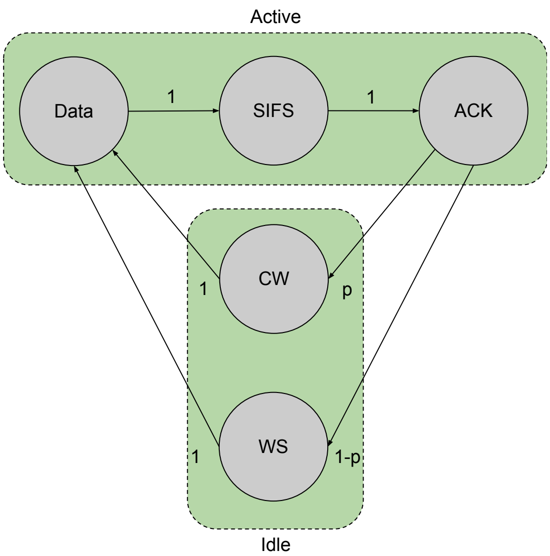


Figure 2.1: WLAN channel model with all the states [4]

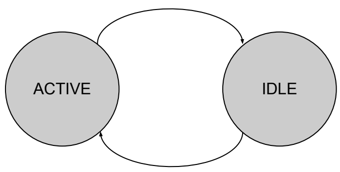


Figure 2.2: Semi-markovian model

The WLAN spectrum activity can be differentiated in two main states [4]. As it is shown in [3], the Data packet, Short Interframe Space (SIFS) and Acknowledge packet (ACK) states are deterministic since the transition probabilities are very close to one and can be merged into a single ACTIVE state. The short duration of the SIFS make this idle period impossible to be used for transmissions. This is why this idle period is merged into the active state. On the other hand, the Contention Window (CW) and the White Space (WS) can be merged into an IDLE state. The different models presented in [4], [7], [10] results in a more simpler model that is represented in Figure 2.2. Since the holding times in each one of both states are not exponential, the continuous Markovian chain properties do not hold. For this, it is necessary to model the holding times in the active and idle states. These holding times can be approximated by two distributions $f_A(t)$ and $f_I(t)$ for the holding times in Active and Idle states respectively. As it is proposed in [4], the active state can be modeled by a Uniform Distribution in the range of α_{on}, β as it is presented in (2.1).

$$f_A(t) \triangleq \begin{cases} 0 & t < \alpha_{on}, \\ \frac{1}{\beta - \alpha_{on}} & \alpha_{bk} \leqslant t \leqslant \beta. \\ 0 & t > \beta \end{cases} \tag{2.1}$$

On the other hand, the idle state is more complex. In [3] the idle state is modeled using only the WS state. For this, since the WS follows a long-tail behavior, a Generalized Pareto Distribution is proposed to model this state. In [4] the model is extended and the idle state is modeled with both CW and WS states, using an Uniform Distribution to model the CW state. The change between states is done with probability p. The final mixture distribution used to model the idle state is expresed in (2.2):

$$f_I(t) \triangleq \begin{cases} p\, f_I^{(CW)}(t) + (1-p)\, f_I^{(WS)} & t \leqslant \alpha_{bk} \\ p\, f_I^{(WS)}(t) & t > \alpha_{bk} \end{cases} \tag{2.2}$$

where the $f_I^{(CW)}(t)$ is an Uniform Distribution in the range of $[0, \alpha_{bk}]$ and $f_I^{(WS)}(t)$ is a Generalized Pareto Distribution to model the WS in $[t > \alpha_{bk}]$. The Probability Density Function (PDF) and Cumulative Density Function (CDF) can be expresed as:

$$g_{[\xi,\sigma]}(t) = \frac{1}{\sigma}\left(1 + \xi\,\frac{t}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)} \tag{2.3a}$$

$$G_{[\xi,\sigma]}(t) = 1 - \left(1 + \xi\,\frac{t}{\sigma}\right)^{-\frac{1}{\xi}} \tag{2.3b}$$

With a location parameter of $\mu = 0$ [10], the final mixture distribution is expressed as:

$$f_I(t) \triangleq \begin{cases} p\,\frac{1}{\alpha_{bk}} + (1-p)\cdot\frac{1}{\sigma}\left(1 + \xi\frac{t}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)} & t \leqslant \alpha_{bk} \\ p\,\frac{1}{\sigma}\left(1 + \xi\frac{t}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)} & t > \alpha_{bk} \end{cases} \tag{2.4}$$

Once we have presented the model for the WLAN spectrum activity is necessary to model the observable load of the sensors. An ideal model where the WSN nodes can observe the whole newtork (Global View [10]) and therefore, all the WLAN spectrum activity will be studied. In addition, we also will take into account the sensing limitation, in which the sensors, due to hardware limitations, just can observe a part of the network and, therefore, just part of the traffic (Local View [10]). It is necessary to differentiate between the model for the spectrum activity, which is represented in Figure 2.2 and these two proposed models (Global View and

Local View) which will model the *observable load* of the WSN sensors. These two 'models' will be presented more extensively in the following sections.

2.1 GLOBAL VIEW MODEL

The Global View model reflects an "ideal" case in which the sensors in the WSN have an unlimited sensing capability. The sensors are able to observe the whole WLAN spectrum activity and, because of that, this observable load can be modelled using the simplified semi-Markovian model presented in Figure 2.2.

From the observed traffic, the sensors must estimate the parameters for the Active (α_{on} and β) and Idle (ξ, σ and p) distributions defined in (2.1) and (2.2) in order to reconstruct the spectrum activity for the prediction process.

This section will present briefly the main characteristics of the Global View model.

2.1.1 Case study

The scenario for this model is represented in Figure 2.3.

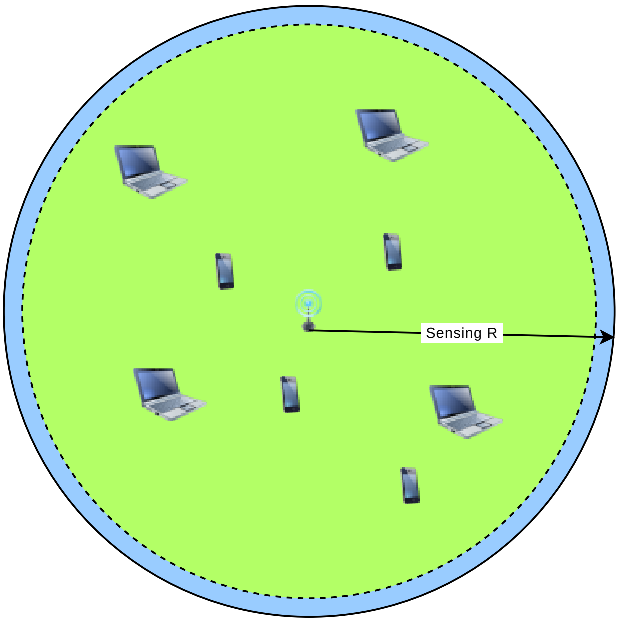


Figure 2.3: Global View model scenario [Example]

The sensing radius of the sensor is the same as the Access Point (AP). The position of the sensor in this scenario is not a critical issue since the unlimited sensing capability. The sensors are able to observe the whole network's traffic.

2.1.2 Estimation

As it has been presented at the beginning of this chapter, the WLAN traffic can be modeled with a semi-Markovian model of two states: active and idle.

The active distribution is composed by two parameters: α_{on} and β. Each one of these parameters can be estimated by the shortest and largest active periods in the network.

On the other hand, the idle distribution presents a more elaborate definition. The distribution is composed by a mixture model which is defined by two distributions: uniform and generalized pareto, defined for $[0, \alpha_{bk}]$ and $[t > \alpha_{bk}]$ respectively. The generalized pareto is defined by (ξ and σ) parameters, while the uniform distribution will be used to determine the third parameter p for the mixture idle distribution. For the estimation of the generalized pareto parameters, the short samples will be filtered and only the larger ones ($t < \alpha_{bk}$) will be used in the estimation process.

The estimation of the ξ and σ parameters for the generalized pareto distribution will be performed using Maximum Likelihood Estimation (MLE). The MLE estimator has been implemented previously in the framework presented in [9] and is a method for estimating the parameters of a statistical model. The method selects values of the model parameters that produce a distribution that gives the greatest probability to the observed data.

Once the estimation of the (ξ and σ) parameters is completed, it is necessary to estimate the p for the mixture distribution in order to delimit the uniform and generalized pareto distributions. This parameter can be obtained in two different ways: MLE or Moment Evaluation (ME). The results presented in [9] show that the MLE is a suitable estimation process for the idle distribution parameters in the Global View model.

A more extensive study of the estimation process for the Global View model has been presented in [10], in which the MLE is presented in detail. In this project we will study deeply the efficiency of the estimation of the Global View parameters in a wider case study of model compliant cases and do the necessary modifications to overcome the possible problems.

2.2 LOCAL VIEW: THE MODEL

While the Global View model presented in Section 2.1 reflects an ideal case in which the sensors have an unlimited sensing capability and are able to observe the whole spectrum activity of the network, in real terms, we need to face a series of issues that can be present in the model. The sensors, due to hardware limitations, have a limited sensing capability, being only able to observe a portion of the network's traffic. So a new model for the observable load of the WSN sensor is needed.

This section will present briefly the Local View model characteristics (scenario and modelling) and estimation process.

2.2.1 Scenario and model

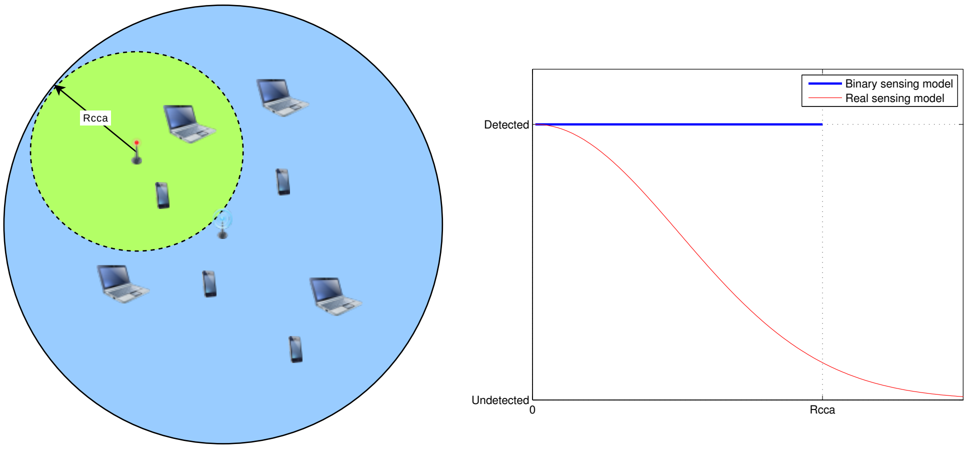
The scenario for the Local View model is represented in Figure 2.4a.

We define a binary sensing model in which the activity of all the WLAN users within the sensor's sensing range will be detected while the ones outside this sensing area will not be observed (Figure 2.4b). However, the real sensing model will decrease its sensing capability gradually. The sensing area of the sensor is denoted as Clear Channel Assessment (CCA) zone.

Because of this limited sensing range, each sensor distributed in the scenario will observe a different portion of the spectrum activity generated by the nodes within its sensing range and therefore, estimate its own parameters from the observed activity. This issue requires an extension of the two-state semi-Markovian model presented in Figure 2.2 in order to represent the unseen activity of the nodes outside the sensing range. The parameters estimated in each one of the sensor can differ depending on the traffic scheme generated by the sensed nodes.

The new semi-Markovian model is composed by three states: idle, active observed and active unseen, represented in Figure 2.5.

It is necessary to define a new parameter that reflects the portion or percentage of traffic



(a) Local View model scenario (example) **(b)** Real and binary sensing models

Figure 2.4: Local View model scenario and sensing model

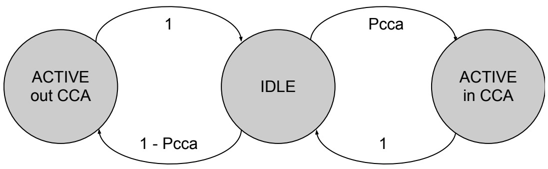


Figure 2.5: Semi-markovian model [10]

seen and unseen. The parameter P_{cca} determines the probability of detected WLAN activity. Consequently, $1 - P_{cca}$ represents the portion of unseen activity.

The active distribution for this new three-state semi-Markovian model is represented by the same active distribution used in the Global View model: uniform distribution. The estimation of the parameters for the active distribution α_{ON} and β_{ON} are estimated by the shortest and largest measured active periods respectively, as it was done in the Global View model. On the other hand, the Global View idle distribution cannot be used in this Local View since the sensor skips a portion of the active periods in the network [10]. For this, the estimation of the parameters cannot be done in the same way as before as it is presented in [10]. The local idle distribution is a combination of the global idle distribution, the active distribution and the observable load P_{cca} and cannot be expressed in a closed form, which do not allow to apply the MLE estimation process presented in [9].

Three different approaches have been presented in [10] and [9] to estimate the idle distribution parameters for the Local View model: compound MLE/ME, Laplace Estimation and Neural Networks. We will focus on the Laplace estimation, which is revised in the following section.

2.2.2 Laplace Estimator

Since the limited sensing capability of the sensors does not allow to apply the idle distribution presented for the Global View model, different solutions have been presented and extensively studied in [10] and [9] for the estimation of the local view idle distribution parameters. Among all of these solutions, the Laplace-based estimation presented a better accuracy in the estimation but with a higher complexity.

Since the Neural Network estimation model has already been implemented and deeply studied in [9], and the Laplace estimator gave a better accuracy in the estimation process, we will focus our study on this second estimator. We will implement the Laplace estimator as an additional tool in the framework implemented in [9].

The Laplace estimator is based in the minimization of the Mean Square Error between both analytic and empirical idle distributions. In this section, we will present briefly the concept and mathematical definitions for the distributions. A detailed presentation of the mathematical basis of the Laplace estimation process is presented in [6].

Algorithm for the Laplace estimator

First of all, we define a discrete state space $\mathbf{K} = K_1, ..., K_K$ where K_n will be a combination of the three parameters that determine the idle distribution: $K_n = (\xi_n, \sigma_n, p_n)$. We want to find the state (combination of parameters) that minimizes the MSE between the analytically determined Laplace Transform and the empirically derived Laplace Transform of the collected samples:

$$(\xi^*, \sigma^*, p^*) = argmin \sum_{S}^{k=0} (f^*_{Ie}(s;N) - f^*_I(s;K_n))^2 \tag{2.5}$$

Both analytic and empirical **LT!** are defined for a finite subset of the Laplace domain "S". In our case, a subset of 1000 "s" samples logarithmically distributed in the range [1,10000]. A higher number of samples implies a better accuracy in the estimation process.

The empirical **LT!** of the observed idle distribution is determined by:

$$f^*_{Ie}(s;N) = \frac{1}{N} \sum_{N}^{i=1} e^{-st_i} \tag{2.6}$$

where, t_i is the idle sample and N is the total number of idle samples gathered for the estimation.

On the other hand, the analytic idle distribution **LT!** is determined by the combination of active and idle distributions and the P_{cca} parameter. The local active distribution follows the same distribution as in the Global View model. On the other hand, the local idle distribution cannot be defined equally because of the skipped activity due to the limited sensing capability of the sensors. The local idle distribution is defined as:

$$f^*_{\bar{I}} = f^*_I(s) \frac{P_{cca}}{1 - (1 - P_{cca}) f^*_I f^*_A} \tag{2.7}$$

where $f^*_{\bar{A}} = f_A$ is the local active distribution, $f^*_{\bar{I}} \neq f_I$ is the local idle distribution and P_{cca} determines the percentage of WLAN activity observed as it has been presented before in Section 2.2.1.

The P_{cca} parameter is defined as:

$$P_{cca} = \frac{\frac{p\alpha_{bk}}{2} + \frac{(1-p)\sigma}{1-\xi} + \frac{\alpha_{on}+\beta}{2}}{\frac{1}{N}\sum_{N}^{i=1} x_i + \frac{\alpha_{on}+\beta}{2}} \tag{2.8}$$

The estimation process is defined in an algorithm already presented in [10] and [6]. The algorithm is presented bellow. In this algorithm, $m \geqslant 0$ denotes the iteration step and $Q_m(K_n)$ denotes the popularity of the state K_n, i.e. the number of times that the algorithm has visited the state until iteration m. K^* denotes the most popular state.

Algorithm 1 Laplace Estimator Algorithm [6]

Step 0:

Choose randomly a starting state $K_0 \epsilon \mathbf{K}$.
$Q_0(K_0) \leftarrow 1$ and $Q_0(K_n) \leftarrow 0, \forall K_n \epsilon K, K_n \neq K_0$.
$m \leftarrow 0$ and
$K^*_m \leftarrow K_0$.
Go to Step 1.

Step 1:

Generate a uniform random variable J_m such that for all $K_n \epsilon K$, $K_n \neq K_0$, $J_m \leftarrow K_n$ with probability $\frac{1}{K-1}$.
Go to Step 2.

Step 2:

Generate an observation R_m of $Z^{K_m \leftarrow J_m}_{lm}$.
if $R_m > 0$ **then**
  $K_{m+1} \leftarrow J_m$.
else
  $K_{m+1} \leftarrow K_m$.
end if Go to Step 3.

Step 3:

$m \leftarrow m+1, Q_m(K_m) \leftarrow Q_{m-1}(K_m) + 1$ and $Q_m(K_n) \leftarrow Q_{m-1}(K_m)$ for all $K_n \neq K_m$.
if $Q_m(K_m) > Q_m(K^*_{m-1})$ **then** $K^*_m \leftarrow K_m$.
else
  $K^*_m \leftarrow K^*_{m-1}$
end if
Go to Step 1.

3 MULTI-LAYER WLAN USAGE MODEL

Our main concern in this project is to test the validity of the presented spectrum activity model against different traffic scenarios. NS2 and its extension NSMiracle provide a simple Constant BitRate (CBR) traffic generator, which will not comply our effort of testing the model against real traffic scenarios. This issue is solved in [9], in which the multi-layer traffic model studied in [8] is implemented and extended with the packet level. This multi-layer traffic model solves our approach of study a real traffic scenario in which validate our Global View and Local View models.

In [8], an extensive study with measurements on real WLAN traffic of a Campus WLAN is developed. This study aims at finding the proper distributions that allow to model the traffic behavior in such scenarios.

The document presents an extensive study of the session and flow levels present in a WLAN network and the distributions that can be used to model both levels. The document present the modelling for multiple access points in the Campus WLAN. We will focus our work on scenarios with a single AP which is also defined in the same document.

In [9] the so-called multi-layer traffic model is implemented with the addition of the packet level. The packet level allows to map the traffic model to spectrum usage. This multi-layer traffic model will give us an insight of how our spectrum occupancy model will behave in this type of scenario.

The traffic for each one of the levels is obtained via randomization following the distributions presented in [8], and that we will present in this chapter. In addition, other distributions had been selected to randomize the traffic for the packet level.

Figure 3.1 represents the traffic hierarchy of the multi-layer traffic model that we are going to use for our simulations.

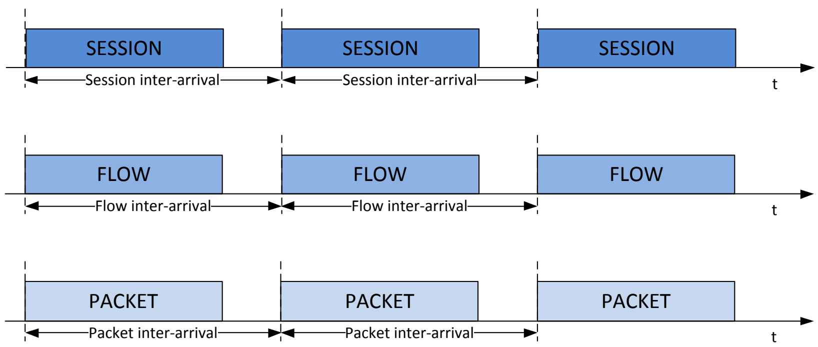


Figure 3.1: Multi-layer WLAN traffic model

As it can be observed, the session level (which represents a user in the WLAN) is composed by the flows and its inter-arrival times. At the same time, each flow comprise of packets with its sizes and its inter-arrivals.

The next sections present a detailed description of the configuration and the distributions to model each one of the levels in the multi-layer traffic model. For a better understanding on how each one of the levels behave, it is necessary to address to the study developed in [8]. An extensive presentation on how to define the traffic in the simulator will be presented later in the chapter dedicated to that topic.

3.1 SESSION LEVEL

The session level is the highest level in this multi-layer traffic model. Each session represents a WLAN user in the network, that connects with an access-point AP. Each session connects to the AP following a time-varying Poisson distribution. This session is composed by a determined number of flows, its inter-arrival times and its sizes.

In [8] it is shown that a bi-Pareto distribution is the best distribution to model the number of flows within a session. On the other hand, the flow inter-arrivals are shown to be log-normal distribution is the best fit.

The parameters that determine both distributions are presented in Table 3.1.

Modeled Variable	Distribution	Parameters
Session arrival	Poisson	$min = 1$, $max = 928$, $median = 11$
Flow number	Bi-Pareto	$\alpha = 0.07$, $\beta = 1.75$, $c = 295.38$, $k = 1$
Flow inter-arrival	Log-normal	$\mu = -1.6355$, $\sigma = 2.6286$

Table 3.1: Session level traffic configuration according to the model in [8].

3.2 FLOW LEVEL

As it has been said, each session consists of a determined number of flows with its inter-arrival times and sizes.

The flow level is characterized by the flow sizes and includes the packet level.

As it is studied in [8], the flow size follows the same type of distribution used for the flow number and can be modeled with a bi-Pareto distribution with the following parameters configuration:

Modeled Variable	Distribution	Parameters
Flow size (bytes)	Bi-Pareto	$\alpha = 0.00$, $\beta = 1.02$, $c = 15.56$, $k = 111$

Table 3.2: Flow size traffic configuration according to the model in [8].

3.3 PACKET LEVEL

Finally, the lower level in this multi-layer traffic model is the packet level. Each flow includes a series of packets and its inter-arrival times. No extended study has been done of this traffic level yet. We will introduce a series of tests to study this level in this project and how it affects to our modeling.

Four different distributions had been implemented to generate the packet sizes and its inter-arrival times: uniform and constant for the packet sizes, and uniform, exponential, log-normal and constant for the inter-arrival times.

3.4 MAC/PHY LAYERS

NS-Miracle and the different modules that compose the software allows to configure the different layers that conform the WLAN communication. The different parameters for the configuration of the physical and MAC layer can be defined in the TCL simulation configuration files provided in the implementation developed in [9].

Each WLAN node is composed by an omni-antenna and transmits with a power of 18 dBm. The different WLAN packets generated by the nodes are transported through the channel and received by the physical layer. It is possible to choose between different propagation models for the packets inside the channel. In our case, following the work developed in [9], the model selected is a simple path-loss propagation model. The transport protocol used for the WLAN communications in our project is TCP/IP.

In our case, as it has been said before, we work with a simplex communication in which the different WLAN nodes deployed in the network will send its packets to the AP.

Once a packet is received by the physical layer, this will decide whether the packet is really observed or not. In a real WLAN, the packet transported through the channel will not be sensed if its power is not enough to be sensed by a WLAN node. In our simulation environment, all the packets can be received by the nodes and then will be filtered. This is done by comparing the power of the received packet with the determined threshold and discarding the packet in case this should not be sensed.

If the packet can be sensed by the node, then it is sent to the next layer: MAC layer. In this layer, the packet is decoded and sent to the proper receiver. The MAC layer takes also control over the shared media as it can be the wireless channel through the media access protocol, detecting and avoiding the re-transmissions and collisions of the packets when accessing to the shared medium. The different parameters for this layer can also be configured through the TCL configuration files provided for the project.

Part II

SIMULATION RESULTS

4 SIMULATION

Previous chapters have introduced the models that are going to be used to model the WLAN traffic behavior. In addition, a short presentation of the traffic model that we are going to use for our experiments, studied in [8] has been addressed. This chapter will introduce the simulation environment and its configuration and will serve as an introduction to the experiments that follow this section.

As it has been introduced previously, the experiments will be run over the Network Simulation 2 software and its NS-Miracle extension [11]. In [9], the implementation of the previously introduced models for modeling the traffic and the multi-layer traffic model have already been developed.

Our work in this project will mainly test the efficiency of these models and its implementation and, if necessary, we will modify/extend the implementations.

In this chapter we will present briefly the simulation environment. For an extended view of the simulation environment see the implementation defined in [9] and the simulation files provided with this project.

4.1 NETWORK SIMULATION 2

NS is a open-source discrete event simulator of networks that provides support for TCP, routing and multicast protocols over wired and wireless networks.

The extension NS-Miracle, is an open-source project that provides a set of libraries of network protocols to fulfill the already implemented in NS2. In addition enables the coexistence of multiple modules in each layer of the protocol stack: multiple IP, link layers, MACs, PHY and so on. NS-Miracle extends the implementation of modern communication systems over NS2. The main library that is going to be used for this project is "dei80211mr", which provides an implementation of the 802.11 protocol over NS2.

In [9] the presented semi-Markovian models and the multi-layer traffic model [8] have been implemented. The protocol stacks for the WLAN and WSN have been also implemented as it is presented in Figure 4.1.

As it can be observed in Figure 4.1, the protocol stack for the WSN is much simpler than the WLAN. The sensor will observe the channel using the physical layer implemented in Module 1. The frame will continue to the next layers if it is above the sensing threshold. Module 2 will process the frame, deciding whether or not is a WSN or WLAN frame, drop, etc. Finally, the third module is the responsible for the estimation process already described in Chapter 2.

4.2 TRAFFIC GENERATOR

As it has been already introduced in Chapter 3, the multi-layer traffic model presented in [8] has been implemented over the NS2 software, more specifically, a C++ library of random distributions defined in Chapter 3. The traffic will be generated randomly following the distributions defined in 3 and using the GSL library.

The bi-Pareto distribution is used in the mixture-idle, for this, is composed by a Uniform

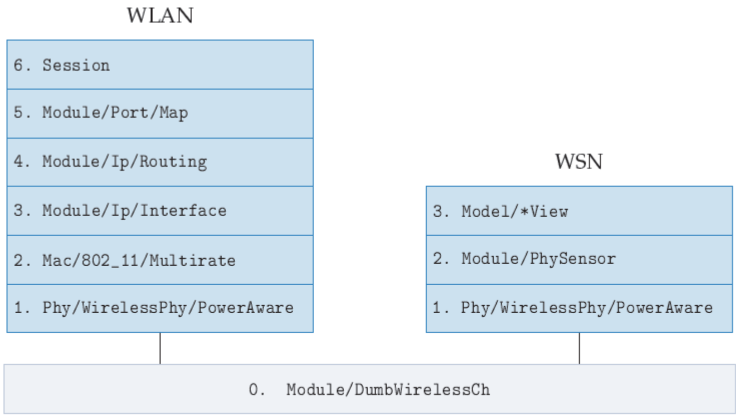


Figure 4.1: Protocol stacks for WLAN and WSN [9]

and a generalized Pareto distributions.

Each one of the distributions can be configured manually choosing the desired parameters via the TCL configuration files provided with the implementation.

4.3 ESTIMATION LIBRARY

The estimation library is implemented in C++ and comprises different modules that define the estimation processes for the Global and Local View models.

The samples are gathered using the "PhySensor" class. This class manages the samples gathered from the network's traffic, decides whether or not is sensed by the sensor in which the PhySensor class is attached using the specified sensing threshold. Decides if the frame is a WLAN or WSN communication and which type in case is a WLAN frame (ACK, **RTS!**, **CTS!**, data, etc.). And finally adds the idle-active duration for the estimation process.

The estimation library is composed by the following modules that will perform the estimation on the idle samples gathered previously by the PhySensor class:

- ModelView: super-class that defines the skeleton for the rest of the estimation modules.
- ModelGlobalView: class that defines the estimation processes for the Global View model. It uses the MLE estimation process. It returns the estimation parameters for the active and idle distributions.
- ModelKSGlobalView: extends the estimation process defined in ModelGlobalView with the Kolmogorov-Smirnov test, which validates the fitting between the empirical and estimated distributions. It returns the D and P values for the K-S test in addition to the active and idle distribution parameters.
- ModelLocalView: class that defines the estimation processes for the Local View model. It uses the Laplace estimator presented in Section 2.2.2. It returns the estimation paramters for the active and idle distributions.
- ModelKSLocalView: same functionality as the ModelKSGlobalView module but for the Local View model.

4.4 VALIDATION TEST

The implementation developed in [9] includes a validation test that will test the fitting between the empirical and estimated distributions.

The validation test will be performed on the truncated part of the mixture idle distribution defined in 2.2. First of all, the validation test will gather the idle samples. For a faster search of the D and P values, only a 10% of the idle samples gathered will be used in the validation test[1].

Secondly, obtain the CDF of the samples and find the maximum deviation between both empirical and estimated distributions. The P-value will be determined by comparing the obtained D-value with the D value from a series of uniformly distributed values.

The pseudo-code for this process is presented below even so, there are other possible methods to obtain the p-value for the K-S validation test.

---

1 Add idle samples only if $t_{sample} > alpha_{BK}$
2 Find the maximum deviation between the empirical and estimated distributions
  for i in $1..num_samples$ **do**
    Find CDF of sample i.
    $d = max(abs(CDF_{empirical} - CDF_{estimated}), d)$
  end for
3 Estimate the P-value.
  for i in $1..num_runs$ **do**
    Generate $N = num_samples$ uniformly distributed values.
    $D = max(abs(CDF_{uniform} - CDF_{estimated}), D)$
   if D >= d **then**
    $count + +$
   end if
  end for
4 Return D and P values.
  $P = count/\#runs$.

4.5 SIMULATION CONFIGURATION

In order to define the simulation environment, a TCL library is provided with the [9] implementation. The simulation environment is started using TCL configuration files in which the different layers of the network and other characteristics are specified.

First of all, the environment configuration is needed. The "init.tcl" configuration file includes a list of the different sub-configuration files that define the characteristics of the network environment.

The traffic generation is defined configuring the random distributions presented in Section 4.2. The simulation time and other parameters as logs, traffic statistics and so on can also be defined.

Then, the WLAN and WSN nodes are deployed uniformly distributed or manually in the scenario. Each one of the WLAN nodes is linked to another node or an access point. These nodes will generate the traffic accordingly to the multi-layer traffic configuration defined.

The sensing capabilities of the WSN nodes need to be defined also in the TCL simulation file. Here, the sensing modules defined in Section 4.3, sensing time, number of samples to gather, etc. are defined.

---

1 In the results chapter, we will see how this decision will affect the K-S performance.

Once the entire configuration is completed, the simulator starts the simulation environment loading the configuration defined in the TCL file. The nodes are deployed in the scenario, the session and flows are initialized and a series of packets will be sent between the nodes during the simulation time defined or until all the flows are empty (for each one of the packets, the size of the flow associated to that packet is decreased).

During this simulation time, the WSN gathers the samples needed for the estimation process defined in the model associated to the WSN node and returns the active and idle distribution parameters.

The following table shows the main configuration files and its characteristics:

Configuration file	**Characteristics**
utility/init.tcl	instantiates the configuration files to be loaded
utility/simulation.tcl	initiate/stop simulator with the parameters defined in the simulation file
config/global.tcl	defines the main parameters and log files
config/trace.tcl	defines which layers will be stored in the trace file
config/layer_N.tcl	defines the parameters for the layer N
config/layer_5_wlan/.tcl	configuration files to define the parameters for the multi-layer traffic generator
classes/*.tcl	initializes the WLAN and WSN nodes (position, sensing model, traffic configuration)

Table 4.1: Main simulation configuration files

5 GLOBAL VIEW: RESULTS

This chapter presents the results of the experiments designed to test the validity of the Global View Model proposed in [7]. In order to test the model, a systematic way has been followed to analyse in a proper way the results.

The experiments carried out in this part of the project consist of a deeper study of the Global View model that was the starting point in [10] with the difference that the tests will be carried on the multi-layer traffic workload model presented in [8] and later extended in [10], in order to test the validity of the proposed model in real WLAN traffic scenarios. In order to perform the tests, we will use the NS Miracle framework and the implementation of the Global View estimator developed in [10]. Different tests will be carried out over the different levels of the multi-layer traffic model. For this, each one of the other levels that are not going to be tested should be fixed for all the runs of the same simulation configuration, changing a single layer at a time, in order to be able to analyse the final results properly. We will focus on the validation of the Global View model for the idle distribution.

The Global View validation process is composed by different phases: the sensors observe the WLAN traffic and estimate the parameters needed in order to reconstruct the idle distribution, finally a validation test is started, which will test the fitting between the empirical idle distribution and the generated from the estimated parameters. We will carry a series of experiments to test each one of the phases and correct possible errors.

5.1 METHODOLOGY

As it has been explained before, the Global View validation can be divided in three phases. We will perform different experiments over each one of these three phases:

- The sensor observe the WLAN traffic and extract the idle distribution.
- The sensor then perform the estimation of the parameters needed in order to reconstruct the idle distribution.
- Finally, a validation test is executed in order to test the fitting between the empirical idle distribution and the one generated from the estimated parameters.

From the results of the experiments we will conclude if each one of the phases are working correctly and, if needed, we will extend and modify the estimation and validation processes to refine the work developed in [10].

5.1.1 Scenario setup

A basic setup will be used for the experiments. Several configurations can be followed to perform the tests. For the experiments developed to test the Global View model, we will define a scenario composed by a determined number of WLAN users uniformly distributed and an Access Point with a coverage of 100 meters. The distribution of the users is not crucial for the Global View model since the sensor is capable to observe the whole traffic. The only condition is to avoid hidden terminals and a WLAN node should be visible by all the other nodes. The

WLAN nodes and AP will use the 802.11 IEEE standard with a transmit rate of 11 Mbps and a transmit power of 15 dBm. The number of users will be determined by the needs of each one of the experiments. In addition, we will deploy a single sensor in the scenario. The sensor can be positioned anywhere since we are working with the Global View model. In our case, we will fix a sensor in the same position as the WLAN AP.

The main characteristic of the Global View model is that the sensor can observe the whole spectrum activity, which means that the sensor will have the same sensing range than the WLAN coverage range. We will use an ideal case in which the sensor will be observing the network from the start of the simulation, considering that all the users already arrived to the network and will gather a determined number of samples that will be used for the estimation process.

5.1.2 Study of the extraction of the Idle Distribution

The first step before the estimation of the parameters is to reconstruct the idle distribution. It is necessary to test if the distribution generated from the idle samples extracted by the sensor, can be modelled using the Idle function defined in (2.2). As it has been presented in Chapter 2, the idle distribution can be approximated by a mixture distribution (see (2.2)) which is formed by a uniform distribution to approximate the CW in the range of $[0 < t < \alpha_{bk}]$, and a pareto distribution for the heavy-tail behavior of WS in $[t > \alpha_{bk}]$. If the idle distribution does not follow this behaviour, the estimation process cannot be done.

Firstly, we will make a visual validation of the idle distribution. We will check that the idle periods generated by the simulator using the multi-layer traffic model can really be approximated by the mixture idle distribution. For this, we will extract the CDF of the idle periods from the trace file generated by the simulator and we should be able to observe the uniform and the heavy-tailed behaviors in the specified areas.

In addition, we will test the effects of active distribution (packet sizes) on the idle behaviour. For this, we will test the sensitivity of the idle distribution against different packetization processes. The packetization process is determined by the packet sizes and their inter-arrival times within a flow, and the WLAN transmit rate. For these experiments we will use different traffic configurations in the packet level, which means that in order to compare the results, we need to fix the values of the other levels of the multi-layer traffic model (session and flow levels). Each of the sessions will use a different number of flows and inter-arrival times.

We will test different active distributions (determined by the packet size) for the same idle distribution (determined by the inter-arrival times between packets) and compare the obtained results. If the process is carried correctly, the results should show that the idle distribution is almost insensitive to the active distribution changes.

5.1.3 Estimation Process

The second part of this chapter will test the estimation process of the proposed model. More in detail, will test the algorithm design and configuration developed in [10]. The estimation process is carried by the sensors which, from the observations over the WLAN spectrum activity, should be able to estimate the needed parameters for the mixture idle distribution in order to reconstruct the WLAN spectrum activity and be able to predict its behavior.

As it is introduced before in this document and in [7] and [10] the estimation process is carried out using Maximum Likelihood Estimation (MLE). The correct functioning of this process is a key point of the model since these parameters will be used by the sensors in order to reconstruct the behavior of the network and predict when it will be idle and hence, be able to send data. The estimated parameters should be the same for the same high-level statistics.

The main goal of these experiments is to test whether or not the randomization of the packet level affects the estimation process, affecting the stability of the estimated parameters that will be used later to reconstruct the idle distribution.

In order to test the MLE, the same simulation with a fixed traffic configuration will be run several times and the estimation parameters will be extracted in order to be studied later. The traffic configuration will be the same as the defined in Section 5.1.2 in which the session and flow levels are fixed while the packet level is randomized. If the estimated parameters do not differ in a high manner, then the estimation process is insensitive to the packet level. We will compare different random distributions in the packet level in order to test the estimation process. From the estimated parameters of the different tests, the mean and standard deviation will be extracted.

5.1.4 Model Validation

Once we have tested the correct functioning of the generation of the idle distribution and the estimation process, is necessary to check if the reconstructed idle distribution fits with the empirical one. For this, we will use a Goodness-of fitness test which is a validation test that will check both distributions and determine how good the proposed model is for the determined scenario. The implemented validation test is the Kolmogorov-Smirnov test. The K-S test is a Goodness-to-fitness test that will test the fitting between the empirical distribution (obtained from the simulated traffic) and the estimated distribution constructed from the estimated parameters. This validation test measures the deviation (D-value) of the empirical and experimental functions and the probability of null-hypothesis (P-value). The scientific community had determined that a null-hypothesis will be rejected if $P - value < 0.05$.

In order to prove that the K-S test is a good Goodness-of-fitness test for the proposed model, different tests will be performed following the same procedure for testing the MLE in Section 5.1.3. Multiple runs of the same configuration will be carried and we will extract the D and P values of the K-S in order to study the deviation between the parameters.

The experiments performed in this section will give us an insight of how optimal is the K-S test for our model and if it is necessary to use another validation test.

5.1.5 Effect of the number of samples in the Kolmogorov-Smirnov validation test

In this experiment we studied the impact of the number of samples in the Kolmogorov-Smirnov validation test. The number of idle samples used for the K-S has an impact in the time of performance. The first implementation of the validation test for our model uses a 10% of the idle samples gathered for the estimation of the idle-distribution parameters. This decision has been made in order to achieve the estimation of the p-value in the K-S test faster. This experiment will give us an insight of the impact of using a percentage of the total set of samples in the performance of the validation test.

5.1.6 Session and in-Session Experiments

Finally, after testing the mixture idle distribution, the estimation process and the validation test, it is necessary to test the combination of MLE and K-S validation test for a wide range of parameterizations of the multi-layer traffic workload model. Instead of full-randomization of the input variables, we will conduct detailed experiments for different areas of each traffic variable [e.g. session/flow number etc.], representing a different "operation points" of the WLAN network. The objective is to test and identify the areas where our modelling fits better and where it fails.

Again, we will test each one of the three phases presented at the beginning of this section with similar procedures followed in each one of the previous sections.

The outcome of the following experiments will be a general evaluation of the designed Global View Model in different traffic scenarios. The results should give insight into possible extensions of the activity model that match better the considered traffic workload model.

5.1.7 Autocorrelation study of the Active periods for the Global View model

As it has been explained in Chapter 2, one of the assumptions in the Local View model is that the different active periods in the network are independent since are generated by independent WLAN users. We will test this assumption. For this, first of all we will study the active periods of the whole network (Global View model), in order to test whether this assumption holds or not. In case the assumption holds, then, in following chapters we will study the independence of the active samples gathered by the sensors in the Local View model, which have a limited sensing range.

Here we will use again the autocorrelation function to study the independence of the active samples. We will use the following set of experiments:

- **Session Experiment**: First of all we will test the independence of the active periods for different number of WLAN sessions. All the cases will have a similar load so we can compare the effect of the number of sessions in the sequence of active periods.
- **In-session Experiment**: Secondly, we will study a single case with a determined number of sessions (i.e. 10 sessions) and different load cases for the same number of sessions. With this, we will study the effect of the load in the active periods sequence.

5.2 RESULTS

In this section we will present the results of the experiments presented in the previous section supporting our conclusions by different graphical representations and tables of values.

5.2.1 Study of the extraction of the Idle Distribution - Results

Mixture Idle Distribution

As it has been explained before, the idle periods distribution is approximated by a mixture idle distribution that is composed by a uniform distribution to model the CW behavior and a truncated pareto to model the WS. We tested the idle distribution generated by the simulator using the proposed multi-layer traffic model.

In order to test this mixture distribution, we generated different tests with a medium-load traffic. The simulation has been done fixing the values of the session (number, number of flows) and flow levels (size, inter-arrival times) and randomizing the packet level (packet size and packet inter-arrival) to configure a medium-load traffic. Different distributions (constant, uniform, exponential, log-normal) have been used for the packet level variables. The user realization for these tests is represented in Figure 5.1.

In [4] it has been explained that the CW follows a uniform behavior that can be approximated by the uniform distribution in the range of in the range of $[0 < t < \alpha_{bk}]$, while the WS follows a heavy-tail behavior that is modelled by a generalized pareto distribution in $[t > \alpha_{bk}]$. Through simulation, we generated a different experiments following the configuration for the WLAN traffic model presented in Chapter 3 and presented in Table 5.1. We extracted the CDF of the

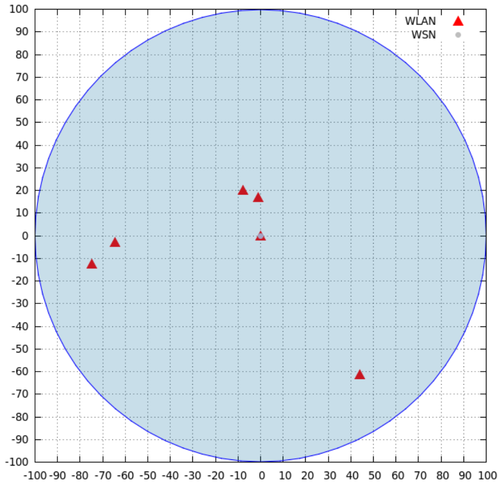


Figure 5.1: Example of a single user realization with uniform placement

duration of the idle periods in order to know if this mixture idle distribution can be used to model the idle periods durations.

Modeled Variable	Distribution	Parameters	
Session number	Fixed	5 users arriving at simulator start	fixed
Flow number	Bi-Pareto	$\alpha = 0.07$, $\beta = 1.75$, $c = 295.38$, $k = 1$	fixed
Flow inter-arrival	Log-normal	$\mu = -1.6355$, $\sigma = 2.6286$	fixed
Flow Size	Bi-Pareto	$\alpha = 0.00$, $\beta = 1.02$, $c = 15.56$, $k = 111$	fixed

Table 5.1: The parameters used for generating traffic according to the model in [8].

Figure 5.2 presents the results of one of the experiments. From the CDF function we can differentiate two clear different areas: one uniform behavior in the range of $[0 < t < \alpha_{bk}]$ and a heavy-tail distribution for $[t > \alpha_{bk}]$. This differentiation in the CDF indicates that is a good fit potential, that means that the idle distribution can be perfectly approximated by the mixture idle distribution defined in (2.2).

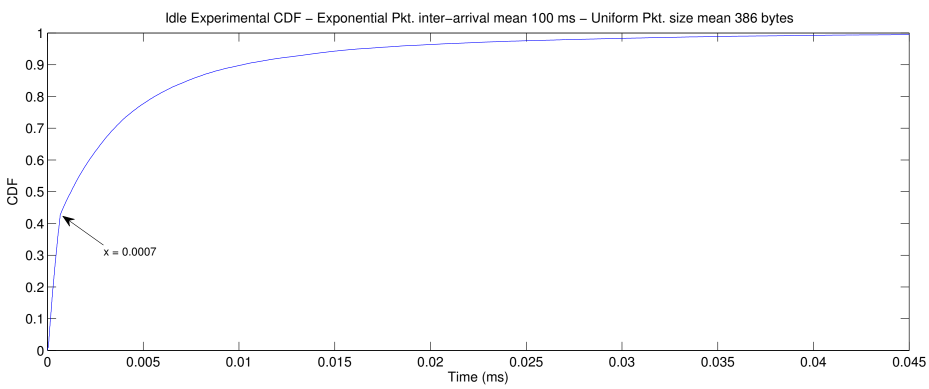


Figure 5.2: Example of an Idle distribution using exponential interarrival time and constant packet size

The same behavior is observed for different distributions for the packet size and packet inter-arrival times.

Idle Distribution Sensitivity

In addition to the experiments developed to visually test the mixture idle distribution, we also tested the idle distribution sensitivity for different active distributions, meaning that, we test if the mixture idle distribution can still be used for different distributions for the packet sizes and inter-arrivals. Again, we generated a medium-load traffic using the traffic configuration presented in Table 5.1 and the same user realization presented in Figure 5.1 and used different distributions for the packet size and packet inter-arrival times and compare the results.

The idle distribution is strongly affected by the packet inter-arrival times while the active distribution is determined by the packet sizes. We fixed one random distribution for the inter-arrival times while using different distributions with the same mean for the packet sizes. In Figure 5.3 we present one example of these experiments. It can be observed that for the same type of distribution for the packet inter-arrival (idle) (i.e. Exponential inter-arrival with 100 ms of mean) and different distributions for the packet sizes (active), the idle periods distribution is almost insensitive to different active distributions.

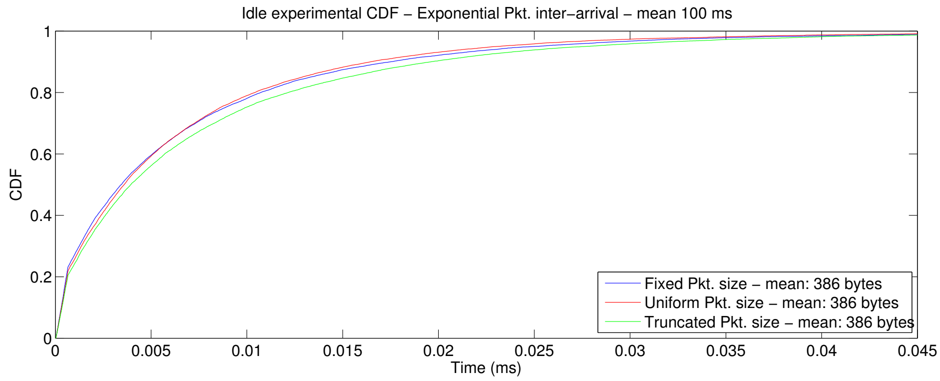


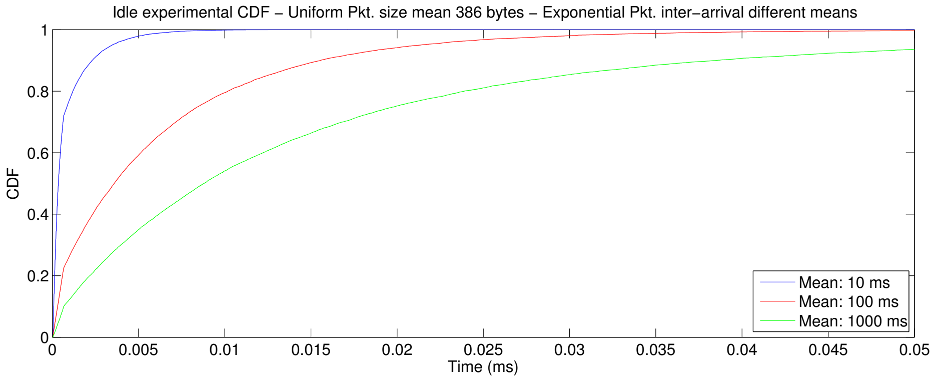
Figure 5.3: Example of different Idle periods distribution for different distributions for the packet sizes. The active distribution do not affect the Idle function behavior.

On the other hand, if the same tests are repeated fixing the distribution for the packet sizes and using different random distributions for the inter-arrival times, it can be observed how different random distributions affect the behavior of the idle function. This is represented in Figure 5.4.

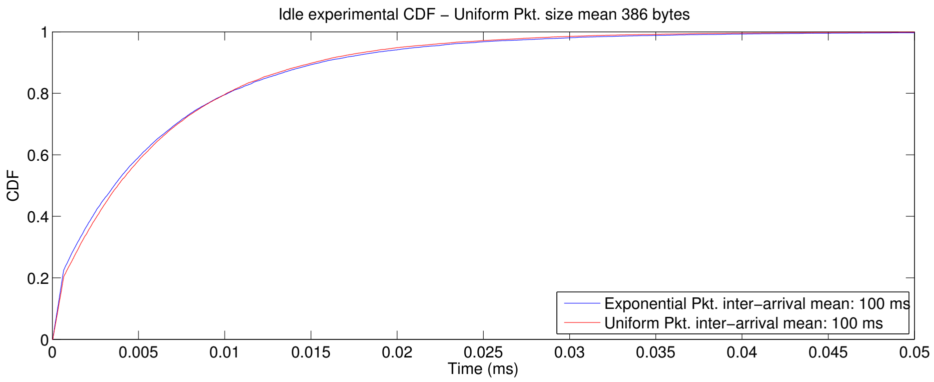
In Figure 5.4a a uniform distribution for the packet size with mean 386 bytes (256 - 512 bytes) has been used for all the tests while we used an exponential distribution with different mean values for the inter-arrival packet time. It can be observed how the proportionality between the mean values of the idle distribution increase/decrease the idle periods distribution represented in the CDF. A low mean value for the packet inter-arrival means shorter idle times and, in extension, higher load in the network, making the packet inter-arrival dominant and reason of this behavior.

In Figure 5.4b is represented a experiment with an uniform distribution for the packet size with mean 386 bytes and different random distributions with the same mean for the inter-arrival times. It can be observed that the idle periods distribution is almost the same for different random distribution if they have the same mean.

Again, the same behavior is observed for different random distributions in the active and idle distributions.



(a) Different mean values for the same packet inter-arrival distribution affect the Idle distribution.



(b) Different packet inter-arrival distributions with same mean do not affect the final Idle periods distribution.

Figure 5.4: Examples of different Idle functions for different Idle distributions

5.2.2 Estimation Process - Results

Once the we proved that the distribution of the idle periods can be modeled using the proposed mixture idle distribution, it is necessary to test the estimation process carried by the sensors. As it has been explained, the estimation process of the parameters for the mixture idle distribution is done using Maximum Likelihood Estimation (MLE). The correct functioning of the estimation process is crucial in order to be able to reconstruct the spectrum activity in the sensors.

We run a medium-load traffic with the simulator using the same configuration for the multi-layer traffic model used in Section 5.2.1, fixing the values for the session and flow levels and randomizing for the packet sizes and the inter-arrival times using different random distributions for each set of runs. We run several times the same configuration of the simulation and extracted the estimated parameters ξ, σ and p. From the extracted parameters we obtained the mean and standard deviation in order to study the deviation between the estimated parameters and decide whether or not the MLE is working properly.

Table 5.2 shows the mean and standard deviation for one of these tests since the results obtained for different random distributions have shown the same behavior.

As it can be observed in the standard deviation in Table 5.2, the variation in the estimation of the parameters of the mixture idle distribution is almost negligible for a same traffic configuration. Since the only level randomized is the packet level, the distribution of the active and idle periods is similar for each run, giving similar estimated parameters, which sustains the proper functioning of the MLE for the estimation of the parameters. From these parameters, the sensors will be able to reconstruct and predict the spectrum activity. We can conclude that

	Mean	Std. Deviation
ξ	0.117043	0.011688
σ	0.00610565	8.42259e-05
p	0.134211	0.00375842

Table 5.2: Estimation parameters statistics - 5 users - Uniform Packet Size (mean: 384 bytes), Exponential Interarrival (mean: 100 ms) - 100 runs

the estimation quality is insensitive to the packetization.

5.2.3 Model validation - Results

As it has been shown in Section 5.2.2, the estimation process is insensitive to the packet process. From the estimated parameters, the sensors should be able to reconstruct the spectrum activity properly using the distributions defined in (2.1) and (2.2).

Once we obtained the reconstructed idle distribution, is necessary to check if it fits with the empirical one (generated from the simulation). For this, we use the Kolmogorov-Smirnov test (K-S) which tests the fitting between both distributions. This validation test will give us a measure of how good the proposed model is for a determined traffic scenario case. Using the same procedure followed in Section 5.2.2, we run several times the same simulation configuration and extracted the D and P values and obtained the mean and standard deviation and the rate of null-hypothesis rejection.

During the experiments developed to test the K-S, different problems had been faced. At the beginning, the test had a high rate of failure, meaning that, that most of the simulations gave us a result of $P - value < 0.05$. This meant that a review of the validation test should be done since the results of the idle distribution and estimation process were the correct ones.

Observing the procedure implemented in [10] of the K-S test in the NS framework, we realized that the test was performed in both uniform and heavy-tail parts of the idle distribution. Due to a high number of the samples gathered by the sensor where in the saturated part ($[0 < t < \alpha_{bk}]$) we decided to apply the K-S test just on the tail of the idle distribution and observe if this correction made the performance of the K-S test better. In addition, the implementation includes just a 10% of the total number of samples gathered by the sensors in order to decrease the execution time of the test. We simulated 200 runs of the same traffic configuration for the session and flow levels, and randomized the packet inter-arrival and size. Each one of the sensors will gather a total of 40000 samples and 4000 will be used for the K-S test. The traffic configuration used for this experiment is presented in Table 5.3.

Modeled Variable	Distribution	Parameters	
Session number	Fixed	5 users arriving at simulator start	fixed
Flow number	Bi-Pareto	$\alpha = 0.07$, $\beta = 1.75$, $c = 295.38$, $k = 1$	fixed
Flow inter-arrival	Log-normal	$\mu = -1.6355$, $\sigma = 2.6286$	fixed
Flow Size	Bi-Pareto	$\alpha = 0.00$, $\beta = 1.02$, $c = 15.56$, $k = 111$	fixed
Packet Size	Uniform	$min = 256$, $max = 512$	random
Packet interarrival (exp. 1)	Exponential	$\lambda = 1$	random
Packet interarrival (exp. 2)	Exponential	$\lambda = 10$	random
Packet interarrival (exp. 3)	Uniform	$min = 0.01$, $max = 0.19$	random

Table 5.3: The parameters used for generating traffic according to the model in [8].

Table 5.4 presents the results of the $P(P - value < 0.05)$ for both the original K-S and the modified K-S tests:

	Original K-S	Modified K-S
Exponential Interarrival (mean: 1 s)	$\approx 4.38\%$	$\approx 5.15\%$
Exponential Interarrival (mean: 100 ms)	$\approx 20.71\%$	$\approx 4\%$
Uniform Interarrival (mean: 100 ms)	$\approx 18.78\%$	$\approx 4.39\%$

Table 5.4: K-S $P(P - value < 0.05)$ - Uniform Packet Size (mean: 386 bytes)

It can be observed clearly that the modification in the K-S test, in which the validation test is just applied on the heavy-tail of the idle distribution gave a better performance reducing the failure rate of the test.

From the experiments we extracted the distribution of the P-value for the different traffic configurations used. This is represented in Figure 5.5.

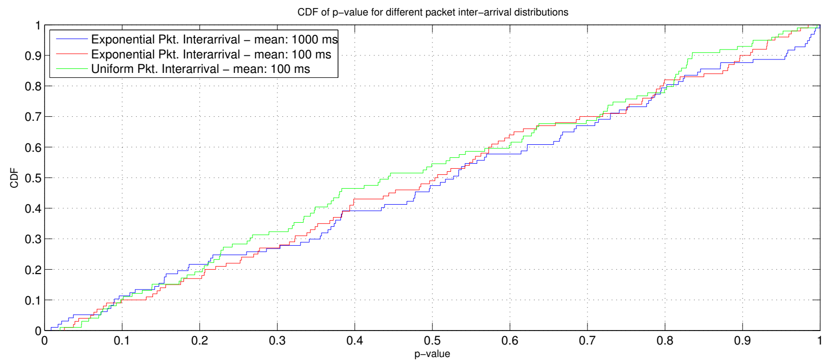


Figure 5.5: CDF of the p-value for different packet inter-arrival distributions (Uniform Packet Size - mean: 386 bytes)

As it can be observed, the distribution of the P-value is similar for the three inter-arrival distributions under study.

For those cases where $P << 0.05$, it is necessary to study if it is still possible to reconstruct the empirical distribution from the estimated parameters. From these cases, we observed that even for a really low P-value, the matching of the empirical and experimental distributions is almost perfect in almost all the cases studied as the example represented in Figure 5.6. This makes us to reconsider if the K-S test is the proper validation test in order to test the validity of the proposed model.

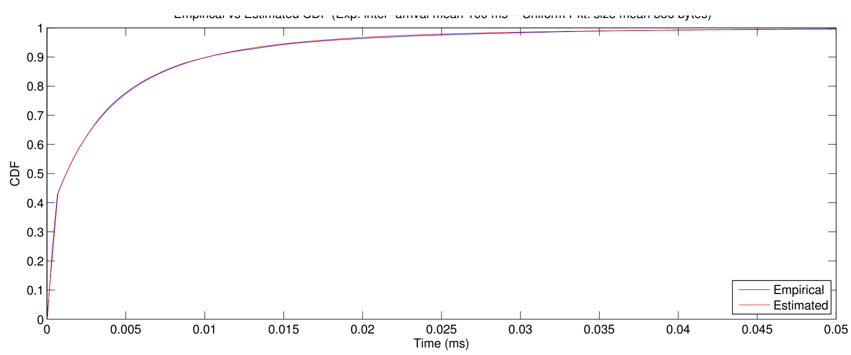


Figure 5.6: Perfect fitting of the CDF of the experimental and empirical functions for very low p-value.

Effect of the number of samples in the Kolmogorov-Smirnov validation test

The first implementation of the K-S test included a 10 % of the total idle samples to perform the validation test. The decision of using a percentage of the total number of samples was done just to perform the validation test faster. In this experiment, we tested the performance of the test using all the samples gathered by the sensors and compared against the first implementation in order to observe if a higher number of samples has any impact in the validation test.

We simulated 100 runs of the same traffic configuration for the session and flow levels, randomizing the packet level as it has been done previously following the configuration in Table 5.3.

We compared the results obtained between both K-S implementations. The results of this experiment are presented in Figure 5.7. We represented the CDF of the p-value of the K-S for 100 runs of each test and compare the performance of the p-value using different number of samples for the validation test. As it can be observed, the performance of the validation test is highly improved using all the samples for the estimation of the p-value, giving a higher value for the P in the K-S test. Table 5.5 includes a resume of the three cases under study in Table 5.4 and compared with the K-S performance using all the samples.

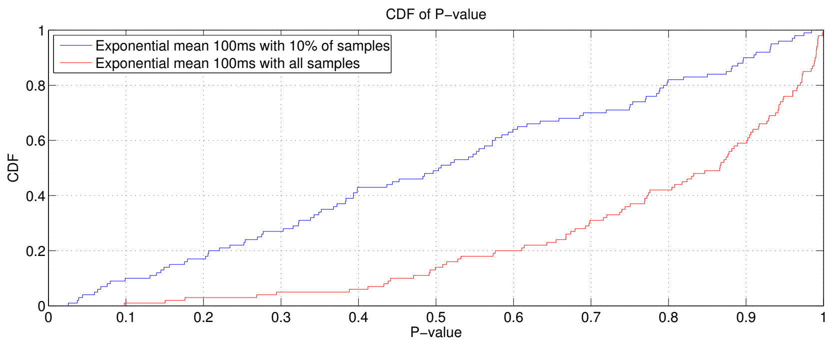


Figure 5.7: P-value failure rate in the K-S test using different number of samples. Exponential packet interarrival with mean 100 ms

	K-S with 10% of samples	K-S with all the samples
Exponential Interarrival (mean: 1000 ms)	$\approx 5.15\%$	$\approx 0\%$
Exponential Interarrival (mean: 100 ms)	$\approx 4\%$	$\approx 0\%$
Uniform Interarrival (mean: 100 ms)	$\approx 4.39\%$	$\approx 3.03\%$

Table 5.5: K-S performance for different packet interarrival distributions and uniform packet size with mean 386 bytes.

As it can be observed from the statistics presented in Table 5.5 the failure rate ($P(P-value < 0.05)$) is clearly improved, and almost null for the exponential inter-arrival cases under study in Section 5.2.3. On the other hand, that improvement is not that clear for the uniform inter-arrival.

Using a percentage of all the samples for the validation test, it is possible that the firsts and tail samples are excluded from the set of samples that will be used to estimate the p-value of the K-S test, affecting the final performance of the validation test. This problem is avoided by using all the set of samples for the validation test and achieving a higher accuracy as it has been demonstrated in this experiment. On the other hand, a higher number of samples implies a higher execution time.

From the results presented in this section, we can conclude that the K-S is the correct validation test to be used for our model if all the samples gathered for the estimation of the parameters are also used for the validation test. Using all the idle samples for the validation test we achieve a better performance and avoid the problems presented at the beginning of this section.

5.3 SESSION AND IN-SESSION EXPERIMENTS - RESULTS

From the results presented in previous sections of this chapter, we can fix now the packet level variables for all the tests, since their effect has been studied. The packet level configuration will be as follows:

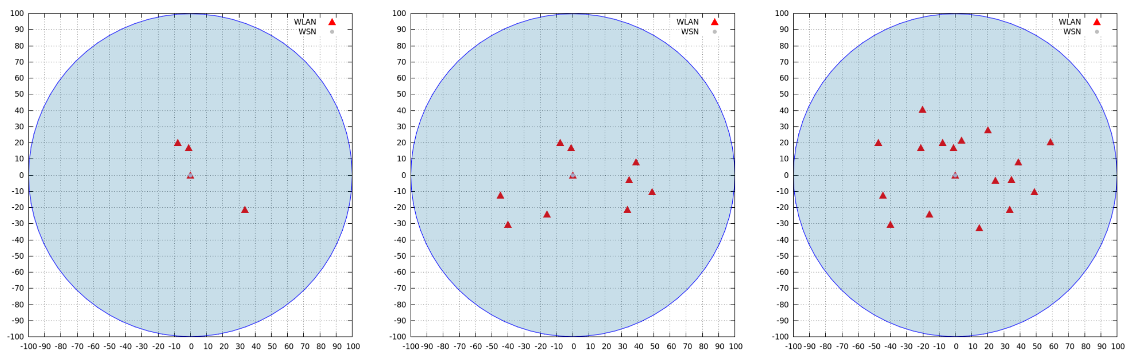
- A uniform packet distribution with an average of 768 bytes [512 - 1024 bytes] for all flows and for all sessions.
- An exponential packet inter-arrival time distribution with some practical mean [e.g. 100 ms] for all flows and for all sessions.

5.3.1 Session Number Experiment

First of all we tested the session level. For this, we used the following parametrization for the session number, e.g. the number of WLAN terminals in the AP area:

- Low session number: **3 sessions**.
- Medium session number: **9 sessions**.
- High session number: **18 sessions**.

The three session cases have a similar load (25%). With this experiment we want to test how the distribution of the load between different users affect the performance of the estimation and validation processes.



(a) 3 sessions (b) 9 sessions (c) 18 sessions

Figure 5.8: Session number experiment - User realizations

All the sessions were configured to arrive at the same time, at the beginning of the simulation, representing that the users had been in the network for a while and then the sensor starts estimating. In real terms, the sensors can be off while the WLAN network is working in order to improve their energy efficiency. The sensors could be activated once some sessions are already in the network transmitting data. For this experiment, the number of flows per session and its inter-arrival times have been fixed in the way that was done before. The packet level has been randomized with the configuration presented at the beginning of this sub-section. We run a total of 25 runs with the same configuration for each one of the three cases. Figure 5.8 represents the user realizations for each one of the three load cases under study in this experiment.

First of all, we need to check if the mixture idle distribution can still be used for the different traffic configurations that are under test. Figure 5.9 represents the CDF of the idle function for the three session cases.

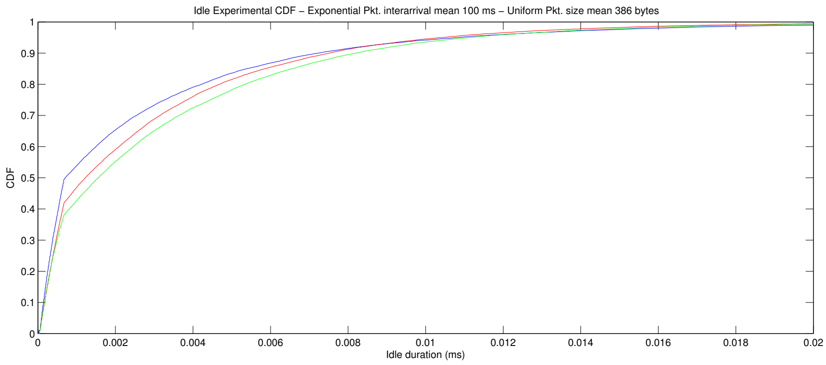


Figure 5.9: Idle distributions for different number of sessions

As it can be observed in Figure 5.9, the mixture model can be used since two differentiated areas can be observed from the idle periods distribution as it has been shown previously.

Since the mixture model can be applied in the three scenarios, it is necessary to check if the estimation process and validation test are being carried properly. For this, we extracted the mean and standard deviation of the estimated parameters and the K-S values for the 25 runs of each case. It can be observed from Table 5.6, that the variability of the estimated parameters and the K-S values can be considered negligible.

	3 sessions		9 sessions		18 sessions	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ξ	0.184869	0.0109214	0.0639	0.0095	0.025	0.0097
σ	0.00337252	4.80725e-05	0.0036	5.9426e-05	0.0045	6.27E-005
p	0.38928	0.00444004	0.2996	0.0057	0.2397	0.004
p_{KS}	0.525856	0 0.253306	0.6874	0.2693	0.796	0.1874
Fail	0 %		0 %		0 %	

Table 5.6: Estimation parameters statistics for different number of sessions

The K-S test shows that there are no fails for the three session-cases under study. The results presented in this section show that the distribution of the load between multiple WLAN users do not affect the estimation and validation processes of the sensors. However, the results cannot be conclusive since are extracted from a 25-runs simulation. A higher numbers of runs are required to have statistically significant results.

5.3.2 In-session statistics

The following experiment randomizes what is happening at the flow level. This level describes the behavior of individual sessions, e.g. WLAN users. For this experiment we tested different number of sessions representing different load regions. Following the traffic model defined in

[8], and the results obtained in this chapter, the different levels of the multi-layer traffic model has been configured as it is presented in Table 5.7. We randomized all the levels simultaneously in order to have multiple possible traffic scenarios in which to test the proposed model. The session cases under study in this these experiment are as follows:

- 5 sessions - Standard traffic configuration.
- 5 sessions - Flow size magnified by 10.
- 10 sessions - Standard traffic configuration.
- 10 sessions - Flow size magnified by 10.
- 15 sessions - Standard traffic configuration.

We decided to magnify the 5 and 10 session cases flow-sizes in order to achieve a higher load range for the same number of sessions but maintaining the distribution shape presented in Table 5.7. This magnification by 10 does not mean that the load will be 10 times higher than the standard configuration since the load is also affected by other parameters in the multi-layer traffic model such as flow number per sessions or packet sizes.

Modeled Variable	Distribution	Parameters
Session number	Fixed	N users arriving at simulator start
Flow inter-arrival	Log-normal	$\mu = -1.6355$, $\sigma = 2.6286$
Flow number	Bi-Pareto	$\alpha = 0.07$, $\beta = 1.75$, $c = 295.38$, $k = 1$
Flow Size	Bi-Pareto	$\alpha = 0.00$, $\beta = 1.02$, $c = 15.56$, $k = 111$
Packet Size	Uniform	$min = 512$, $max = 1024$
Packet inter-arrival	Exponential	$\lambda = 10$

Table 5.7: The parameters used for generating traffic according to the model in [8].

We run 250 runs for each one of the five sub-experiments. For each run we extracted the D-value and P-value from the KS test and the load. The session level (number of sessions and inter-arrivals) will be fixed for all the runs of each sub-experiment while the flow and packet levels will be randomized following the distributions presented in Table 5.7.

Figure 5.10 presents the CDF of the loads for each one of the five experiments under study in this section. As it can be observed from the figure, the load is increased with the number of sessions. At the same time, for a same session case, the load also increases when the flow size is magnified.

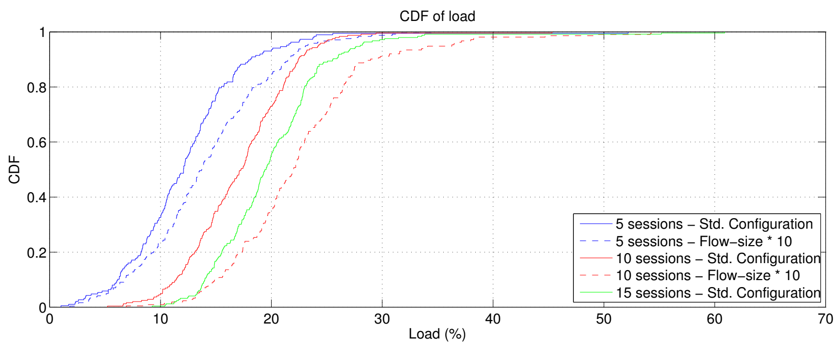


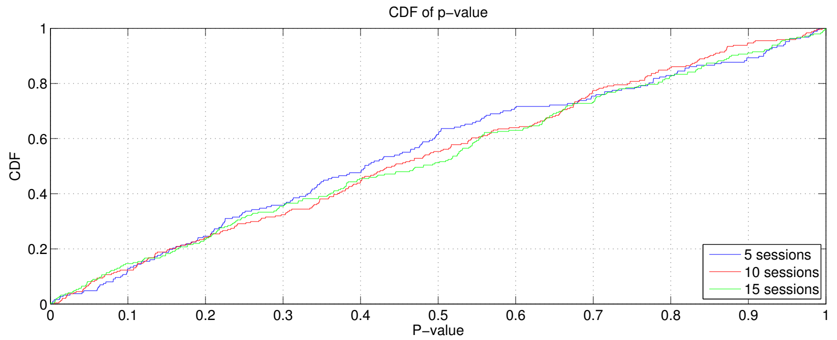
Figure 5.10: CDF of the loads for each one of the five experiments

Experiment	Load
5 sessions - Standard Configuration	12.202 %
5 sessions - Flow size * 10	14.0544 %
10 sessions - Standard Configuration	17.1154 %
10 sessions - Flow size * 10	22.4922 %
15 sessions - Standard Configuration	19.9669 %

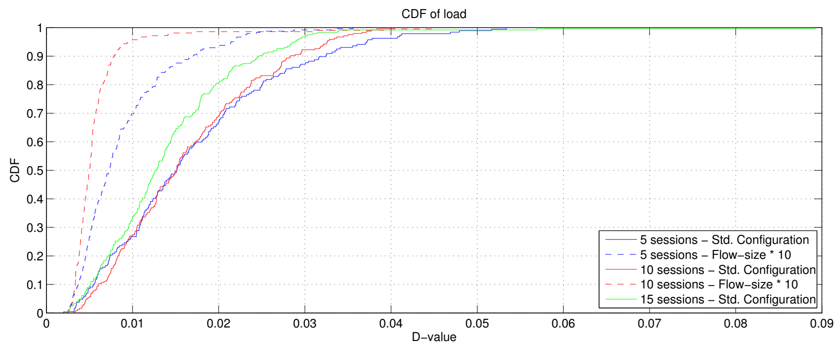
Table 5.8: Average load for each one of the sub-experiments

It can be observed that the load for all the experiments is concentrated around the 15 - 25 % range. The mean load value for each experiment is presented in Table 5.8.

We represented the CDF of the d and p values from the KS test. This is represented in Figures 5.11. As it can be observed the P-value is almost equally distributed for all the experiments (Figure 5.11a), with a failure rate of around 5 % for all the cases. On the other hand, in Figure 5.11b it can be observed that the D-value is clearly improved when the load is higher.



(a) P-value



(b) D-value

Figure 5.11: CDF of P-value and D-value from the KS test for each experiment

Our main goal in this project is to determine when our model can be used. In order to have a better insight on how the model behaves, we filtered the results in different load regions to study how the p and d values are affected. Figure 5.12 represents the number of tests for each region (aggregate from all the five experiments). As it can be observed, a big amount of the tests are concentrated in the region of 10 to 25 % load region (this also can be observed in Figure 5.10).

We represented the mean p value for each one of the load regions for each sub-experiment in Figure 5.13. The most significant results are represented in the load region where all the tests

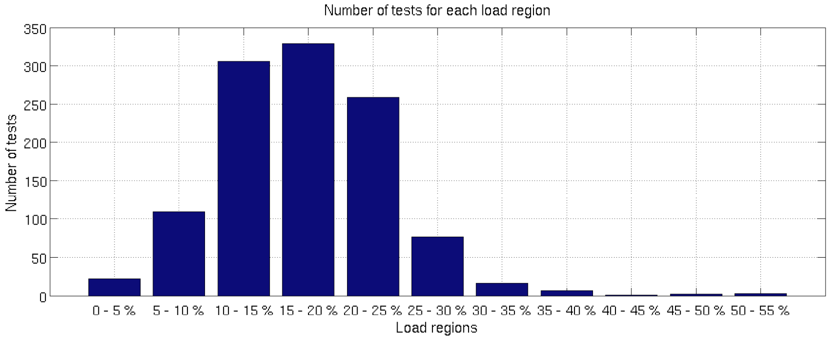
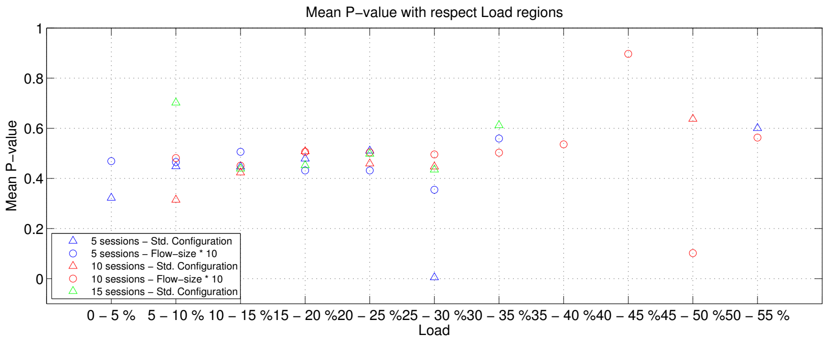
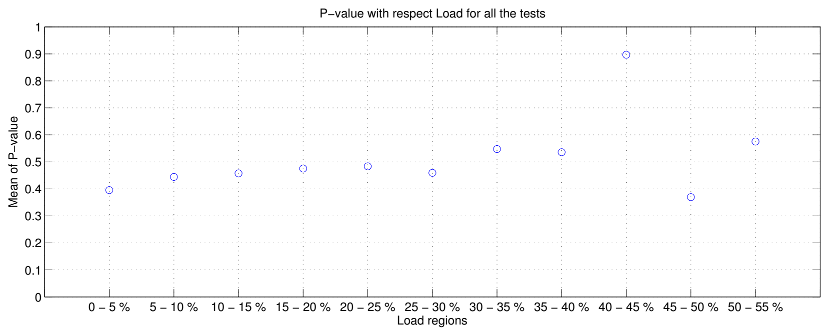


Figure 5.12: Number of tests per load region (aggregate from all the experiments)

are concentrated (from 10 to 25 % of load). In this region, the mean p-value is almost the same for the five experiments, which means that it is not affected by the number of sessions. The rest of the regions represent a small amount of tests, this is why the difference between the p-values is really high between the experiments. In order to have a higher number of tests in each region, we plotted the aggregate mean p-value for all the tests (Figure 5.14b). It can be observed how the p-value is approximately the same for each one of the regions, even so, we may need a higher number of runs of each experiment to have a better conclusion.



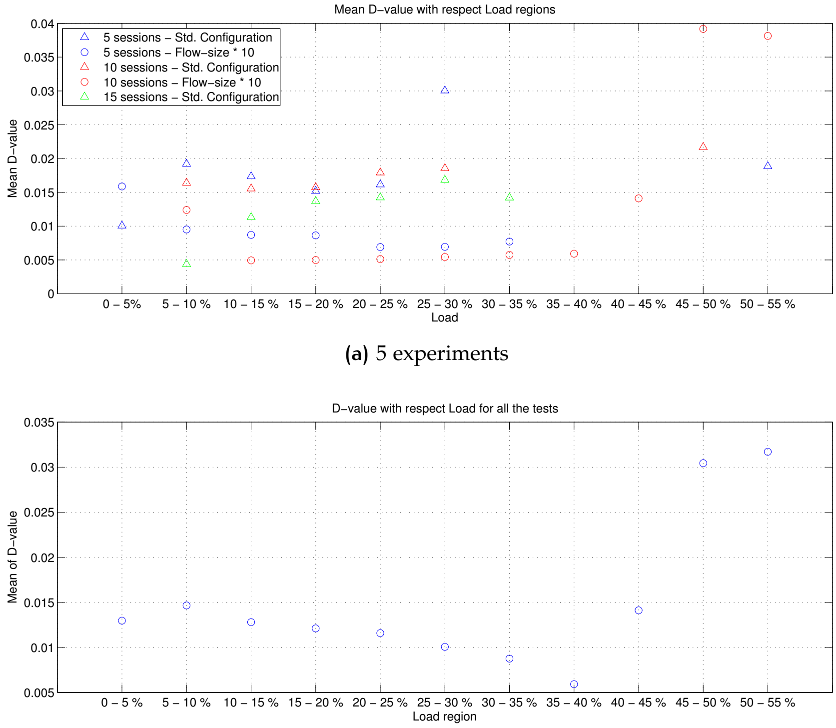
(a) 5 experiments



(b) Aggregate

Figure 5.13: Mean P-value for the different load regions

We did the same for the d value as it is represented in Figure 5.14. In this case, we have a similar behaviour as the presented in the p-value results. The variability between experiments is low enough to be considered null in the load regions with higher concentration of tests.



(a) 5 experiments

(b) Aggregate

Figure 5.14: Mean D-value for the different load regions

From the results presented, we can conclude that our model can be applied for different session cases in the region of 10 - 30 % of load, achieving a low failure rate in the KS test and similar results between different number of sessions. On the other hand, an extended study of lower and higher load regions is needed in order to determine how our model works in those cases.

5.3.3 Autocorrelation study of the time sequence of sources of activity in the WLAN - Results

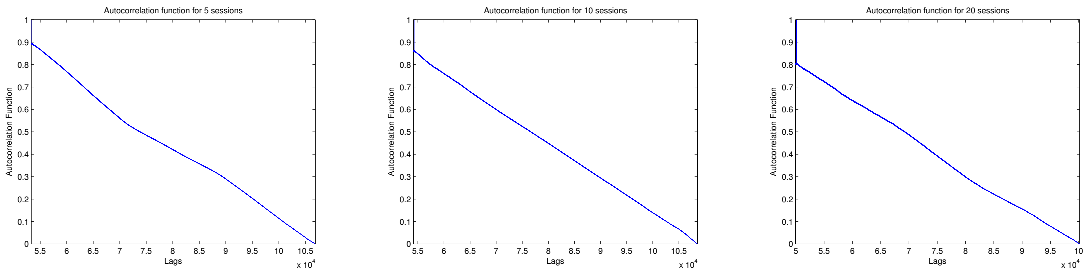
Session Experiment - Results

This experiment will study the session level influence in the independence of the active samples. Following a similar procedure as the one followed in Section 5.3.1, we used different number of sessions with similar load and studied the autocorrelation function of the active samples of each one of the cases. The flow level (size, inter-arrival) configuration is fixed and we randomized the packet level with the following configuration:

- Packet Size distribution: A uniform packet distribution with a normal average [e.g. 768 bytes] for all flows and for all sessions.
- An exponential packet inter-arrival time distribution with some practical mean [e.g. 100 ms] for all flows and for all sessions.

The final configuration for the different levels of the Extended Multi-Layer traffic model is presented in Table 5.7. We selected three session cases, i.e. 5, 10 and 20 sessions with a load of around 20 %.

The independence of the active samples is studied using the autocorrelation function. The autocorrelation function will determine the correlation of the sequence of samples and will give us an evaluation on how independent are the consecutive active periods. We have extracted the sequence of active periods (just focusing on the DATA packets) from the traffic generated with the extended multi-layer traffic model which contains which stations (WLAN users) had sent the active period. We have obtained the autocorrelation from this sequence of stations. These results are represented in Figure 5.15.



(a) (b) (c)

Figure 5.15: Autocorrelation function of the different sources of DATA packets for different number of sessions with similar load. The correlation between samples is reduced for higher number of sources.

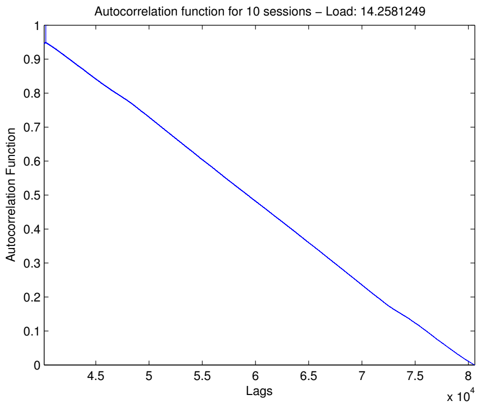
In Figure 5.15 we plotted the autocorrelation function for three different number of sessions cases and similar load in order to compare them. An independent sequence is represented by an autocorrelation function that is approximated by a δ. As it can be observed, there is an improvement in the independence of the active periods when the number of sessions is higher. It is clear that the higher is the number of sessions, more random will be the access to the medium of the different stations of in the network. It is possible to conclude from this experiment that the session number decreases the correlation between the samples. However it is not possible to conclude that consecutive active periods are totally independent as it has been assumed in the Local View model.

In-Session Experiment - Results

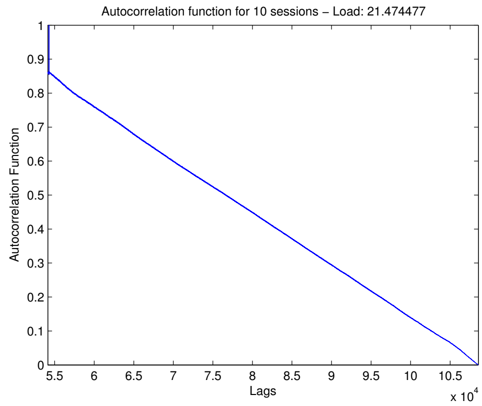
Once we tested the effect of the session number in the independence of the sequence of active samples, it is necessary to study how the load affects this independence. In this case, we repeated the experiments using a determined number of sessions (10 WLAN users) and we randomized the flow and packet levels and extracted again the autocorrelation function from the stations sequence. The configuration of the different levels of the extended multi-layer traffic model can be found in Table 5.7 and the packet level is configured as it was done in the previous experiment. The results of this experiment are represented in Figure 5.16.

From the results of this experiment, it can be observed that the load has a slight effect in the autocorrelation function and, in extension, the independence of the samples. It can be observed how the autocorrelation function tends to a delta function for higher loads. It can be concluded then, that the increase in the load is translated in a higher independence of the samples. When the network is saturated this independence is almost perfect because of the high random access of the packets to the medium. However, the improvement is not that critical as the one presented in the Session experiment.

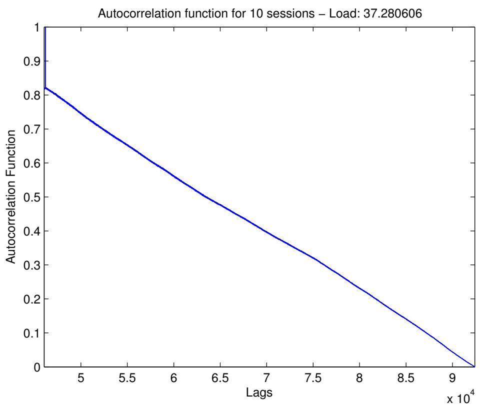
With the autocorrelation experiments developed in this section, we have shown that the active samples can be considered independent enough, so the assumption in the semi-Markovian



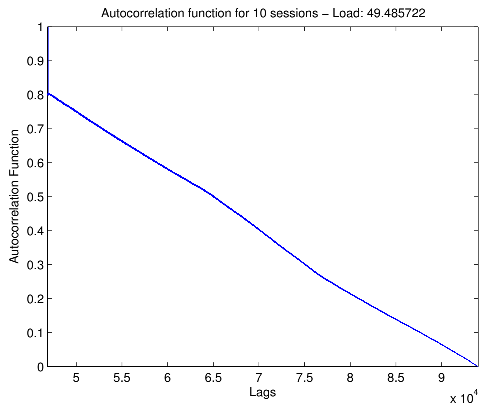
(a) 14.2581249 %



(b) 21.474477 %



(c) 37.280606 %



(d) 49.485722 %

Figure 5.16: Autocorrelation function for 10 sessions for different load cases

model that consecutive active periods are independent has been proved and we will continue with the study of the Local View model in the next chapter.

6 LOCAL VIEW: RESULTS

As it has been explained in previous chapters (see Chapter 2.2), the sensors have a limited sensing capability due to hardware limitations. In the Global View case, we studied an ideal case in which the sensors can observe the activity of the whole network. On the other hand, the Local View case, the sensing limitation is translated into a limited observation of the spectrum activity of the network: the sensors will only be able to observe the traffic generated by the WLAN users inside the sensor's sensing range.

In order to solve this, a three-state Semi-Markovian model has been proposed in [10] (represented in Figure 2.5) in which the observable and non-observable active periods are taken into account for the model using the P_{cca} probability, which determines the amount of non-observable traffic. The assumption in this model is the fact that the consecutive active WLAN periods are independent, i.e. are originated from independent WLAN users or "sessions". This will be the starting point for the experiments for the Local View model.

Chapter 5 introduced the results of the different experiments developed for the Global View model, including a full set of experiments for the next points:

- Testing of the fitting of the semi-Markovian model for the Extended Multi-Layer traffic model proposed in [8] and [10].
- Testing of the estimation process for the parameters: testing the MLE process and the construction of the estimated distribution using the parameters.
- Validation test of the fitting of the estimated distribution against the empirical using a Goodness-of-fit test: Kolmogorov-Smirnov test.
- Study of the Kolmogorov-Smirnov test.

This chapter will present a set of experiments and its results for the Local View Model. Following the same procedure developed for the Global View model, we will test the fitting of the 3-state semi-Markovian model for the Local View model, the estimation process and the validation test for the fitting between empirical and estimated distributions.

6.1 METHODOLOGY

he Local View model can be divided in three phases. We will perform different experiments over each one of these three phases:

- The sensor observe the WLAN traffic and extract the idle distribution.
- The sensor then performs the estimation parameters needed in order to reconstruct the idle distribution.
- Finally, a validation test is executed in order to test the fitting between the empirical idle distribution and the generated from the estimated parameters.

From the results of the experiments we will see if each one of the phases are working correctly and, if needed, we will extend and modify the implementation to refine the work developed in [10].

6.1.1 Scenario Setup

A basic setup will be used for the experiments. Several configurations can be followed to perform the tests. For the experiments developed to test the Local View model, we will define a scenario composed by a determined number of WLAN users uniformly distributed and an Access Point with a coverage of 100 meters. The WLAN nodes and AP will use the 802.11 IEEE standard with a transmit rate of 11 Mbps and a transmit power of 15 dBm. The number of users will be determined by the needs of each one of the experiments. In addition, we will fix different number of sessions depending on the needs of the experiments.

The sensors in this case, have a limited sensing range due to hardware limitations. In difference with the Global View model, in which the sensors are capable of observe the whole traffic of the network, in the Local View model, the sensors can only observe the traffic generated by the WLAN users that are within the sensor's sensing range. This makes more difficult to reconstruct the spectrum activity since the sensors don't know what is happening in the rest of the network.

6.1.2 Active Distribution

In this section we will study the independence of the active periods in order to investigate the assumption in the 3-state semi-Markovian model, in which is stated that the consecutive active WLAN periods are independent, i.e. are originated from independent WLAN users or 'sessions'.

We will study the autocorrelation sequence of the active periods of different traffic configurations of the Multi-layer traffic model and see whether the assumption of independence holds using the same procedures developed for the experiments of the Global View model, where we followed a systematic way when designing the tests.

In addition, we will study the sequence of consecutive skipped active periods due to the limited sensing range capabilities of the sensor. It has been proved that this sequence follows a geometric distribution. For this, we will test the fitting between the empirical and geometric distributions.

It is known that the Kolmogorov-Smirnov test is not a proper validation test for discrete distributions. For this, other alternatives such as Mean-Square Error and Chi-square validation will be tested.

This section will present the results of the independence study as an introduction to the experiments for the rest of the Local View experiments.

6.1.3 Session and Load Experiments

Once the active periods independence and the consecutive skipped active periods has been studied, we will proceed with the final set of experiments for the Local View model.

For this section, a series of tests over the session level has been developed. Here, we will test the validity of the Local View model over different session and load cases, and how the estimation process is carried out.

These experiments will give us a validation of the results previously presented.

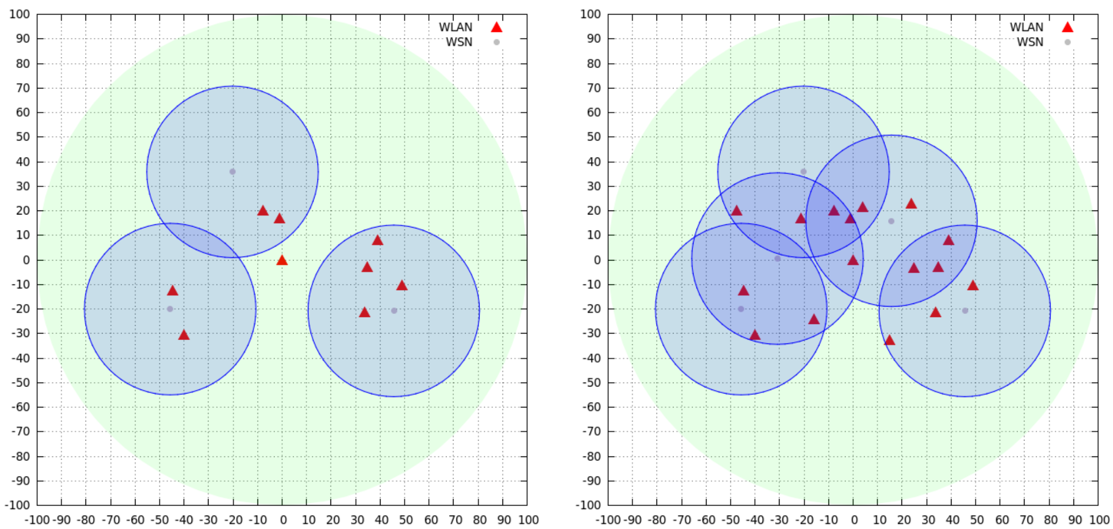
6.2 RESULTS

This section will present the results for the different experiments developed to test the proposed Local View model.

6.2.1 Scenario setup

In this experiment, we placed different sensors in the network. The sensors have a limited sensing range, which means that they are only able to observe a portion of the total traffic. We have studied the independence of the samples gathered by the sensors in order to know if it is possible to use the semi-Markovian model proposed for the Local View model.

We used two different user realizations and number of sensors. This is represented in Figures 6.1a and 6.1b.



(a) 8 sessions

(b) 15 sessions

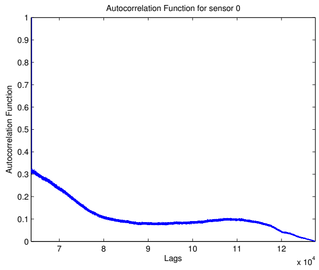
Figure 6.1: User realizations for the Local View autocorrelation experiments (examples)

As it can be observed, each one of the sensors observe a different amount of WLAN users. This will give us the opportunity to study how different number of sessions and load affect the correlation of the active periods gathered by each one of the sensors. We did not performed the estimation process for this experiments.

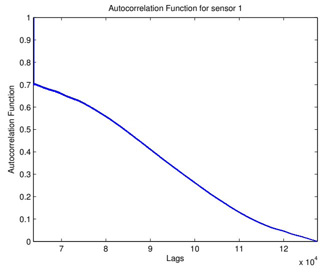
6.2.2 Autocorrelation of the IN/OUT sequence of DATA packets in Local View - Results

From the simulation trace files, we obtain the sequence of active periods and which WLAN stations are generating these periods. In order to study the autocorrelation of the samples, we obtained which WLAN stations are observed by each one of the sensors and, from the stations sequence obtained from the simulation's trace file, we generated a IN/OUT sequence depending on whether the stations are observed by the sensor or not. We finally obtained the autocorrelation function of this IN/OUT sequence. The results are presented in Figure 6.2 and 6.3.

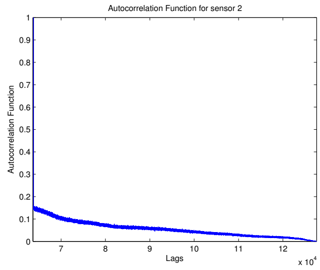
As it can be observed in the 8-session case, the autocorrelation property of the sequence for the two of the sensors (0 and 2) can be considered good: the harmonics are almost negligible and the function is approximated to a δ. This means that the samples in this case are independent and this case is a good candidate in order to apply the Local View model since the assumption of independence of the active periods holds. It can be observed how the autocorrelation function in Sensor 1 is slightly worst.



(a) Sensor 0 - Obs: 2 sessions

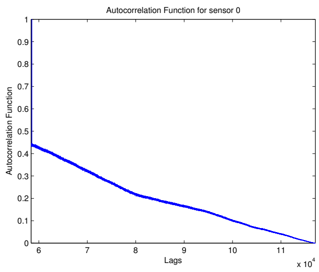


(b) Sensor 1 - Obs: 4 sessions

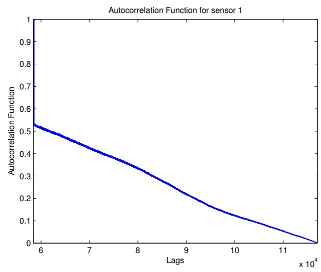


(c) Sensor 2 - Obs: 2 sessions

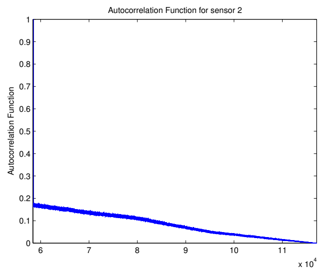
Figure 6.2: Autocorrelation of the IN/OUT sequences for 8 sessions.



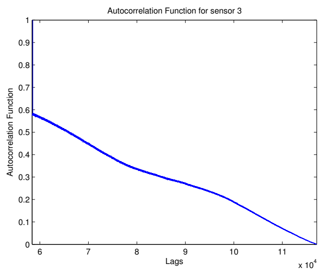
(a) Sensor 0 - Obs: 5 sessions



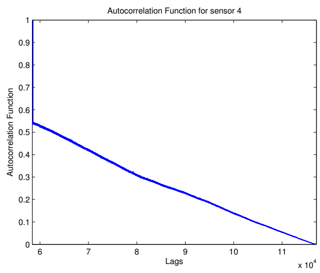
(b) Sensor 1 - Obs: 7 sessions



(c) Sensor 2 - Obs: 3 sessions



(d) Sensor 3 - Obs: 8 sessions



(e) Sensor 4 - Obs: 8 sessions

Figure 6.3: Autocorrelation of the IN/OUT sequences for 15 sessions.

For those cases in which the sequence of consecutive active periods is non-correlated (Sensors 0 and 2 in the 8-session case, for example), the Local View model can be applied properly since the assumption of correlation holds. However, we are facing an important problem: if the sensors need to study the correlation of consecutive active periods from the IN/OUT sequence in order to determine if the Local View can be applied, they need to have knowledge of how is the spectrum activity outside the sensor's range, which means that the sensors need to have a Global View capability or cooperate with other sensors in the WSN in order to determine this IN/OUT sequence.

6.2.3 Density study of the consecutive skipped Active periods in Local View - Results

In extension to the autocorrelation study of the active periods in Section 6.2.2, we will study the correlation of consecutive skipped active periods using the autocorrelation function. For this, we will construct a sequence of number of consecutive active periods that the sensors are unable to observe because of WLAN users outside sensing range.

The traffic of the network is formed by a sequence of active-idle periods generated by the different users. A sensor with a limited sensing range will only observe a portion of this traffic, generated by the users within its sensing range. Therefore, the sensor will 'miss' some of the active periods that are generated by users outside its sensing range. Using the same users realizations used in Section 6.2.2, we will extract the sequence of the consecutive skipped active periods and perform an autocorrelation study in order to observe its behavior.

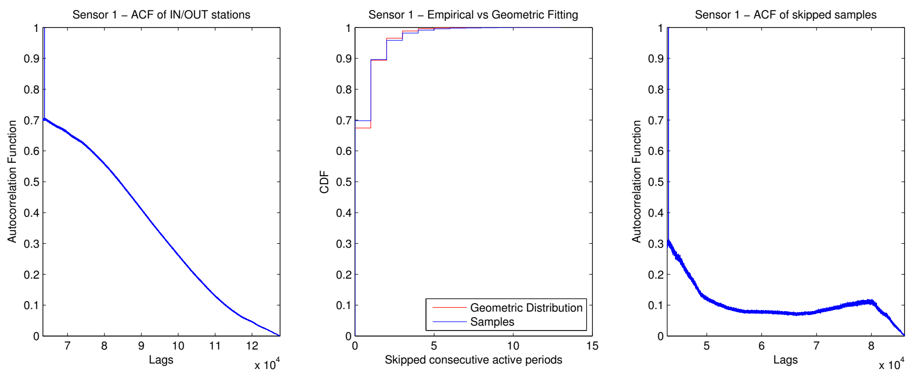
From the first results obtained in this experiment, we observed that the sequence consecutive active periods follow a Geometric Distribution. The geometric distribution is a memory-less distribution and is characterized by a single parameter: p. This parameter can be obtained from the mean value of the samples. For this, we obtain the mean value of the sequence of consecutive skipped active periods of each sensor and obtain the p value by:

$$p = \frac{1}{1 + E(x)} \tag{6.1}$$

With this parameter we will construct a geometric distribution and we will test its fitting with the empirical distribution from the samples. The CDF of the geometric distribution can be found with:

$$CDF(k) = 1 - (1 - p)^{k+1} \tag{6.2}$$

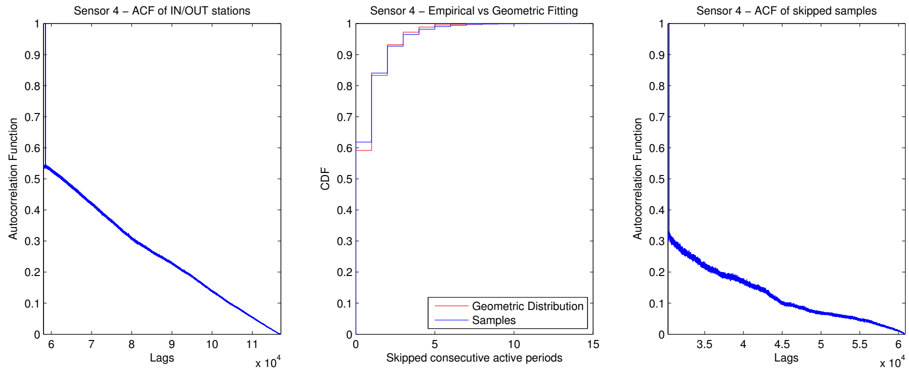
The results obtained from this experiment can be observed in Figures 6.4 and 6.5. The user realizations and number of sensors are represented in Figure 6.1.



(a) Sensor 0: Observes 2 users

Figure 6.4: 8 sessions - Example of fitting of the geometric distribution and autocorrelation of the IN/OUT sequence + consecutive skipped active periods

As it can be observed from Figures 6.4 and 6.5, we have represented the autocorrelation function of the IN/OUT sequence for the sensor (see Section 6.2.2), the fitting of the empirical and geometric distributions, and the autocorrelation function of the sequence of consecutive skipped active periods. From the autocorrelation function of the skipped samples we can say that there is independence between consecutive skipped samples, which was already expected from the results presented previously. In addition, we also represented the fitting of the CDF of the skipped samples and its approximation using the Geometric Distribution. As it can be observed from this fitting, it can be said that the geometric function is a good candidate in order to model the consecutive skipped samples for the Local View model.



(a) Sensor 4: Observes 8 users

Figure 6.5: 15 sessions - Example of fitting of the geometric distribution and autocorrelation of the IN/OUT sequence + consecutive skipped active periods

Goodness-of-fit test for the consecutive skipped active periods

As it has been done in previous sections, the fitting can be tested using a Goodness-of-fit test. In this case, the Kolmogorov-Smirnov test cannot be used. The K-S test requires the samples to be independent and follow a continuous behavior. Our sequence of skipped samples follow is represented by a discrete distribution. For this, the K-S cannot be used to test the fitting of the geometric distribution against our empirical data.

First we studied a modification of the Kolmogorov-Smirnov test for discrete distributions [1], but it was not the easiest solution to implement. We proposed and tested two different solutions to test the fitting which were more easy to implement with the available data: Mean Square Error test and Chi-squared Goodness-of-fit test.

The Mean Square Error (MSE) is based on testing the deviation between both distributions (empirical and geometric) and obtaining the mean square error using the expression (6.3), where N determines the number of samples for the validation test and k is the $k_t h$ sample.

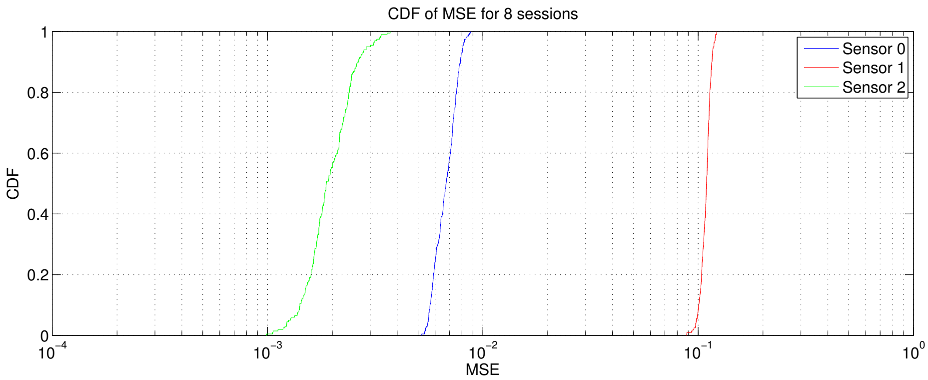
$$\sum_{k=0}^{N} [P_{geometric}(k) - P_{empirical}(k)]^2 \tag{6.3}$$

For this test, we run 100 times the same simulation configuration and extracted the MSE between both distributions for each run. The results are represented in Figure 6.6 in which we have represented the CDF of the MSE for the 100 runs.

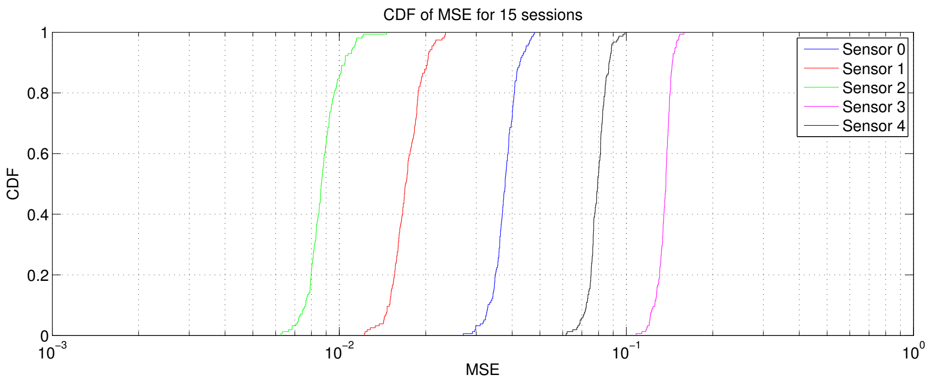
On the other hand, we also tested the fitness of the empirical and geometric distributions using the Goodness-of-fit test Chi-squared. This test can be applied for both continuous and discrete distributions. For this test we also run 100 times the same simulation configuration and extracted the CDF of the p-value. The Chi-square test has been obtained using the already implemented function in MATLAB. The results are presented in Figure 6.7.

For the Chi-square test, as it is done in the Kolmogorov-Smirnov, the confidence limit to reject the null hypothesis is 0.1. In this case, we can observe that there is a high percentage of the 100 runs that the null hypothesis is rejected. In order to know if this failure rate is because of the MATLAB implementation of the Chi-square test which can be to restrictive or because of bad fitting, we plotted the fitting of the empirical and geometric distribution for a high and low p-value of the validation test. This is represented in Figure 6.8.

As it can be observed in Figure 6.8a, where the p-value of the validation test is really low, the fitting between both distributions is clearly bad and in extension the validation test is failing. On the other hand, Figure 6.8b represents one of the best p-value cases for one of the sensors and it can be observed how the fitting between distributions is almost perfect. However, this behavior was also present in those fail cases in which the p-value was close to the confidence limit $p < 0.1$. The high failure rate of the Chi-squared test may be caused because the implementation in MATLAB is too restrictive and the minimum deviation between distributions affects highly the final p-value. A further study in this topic is needed.

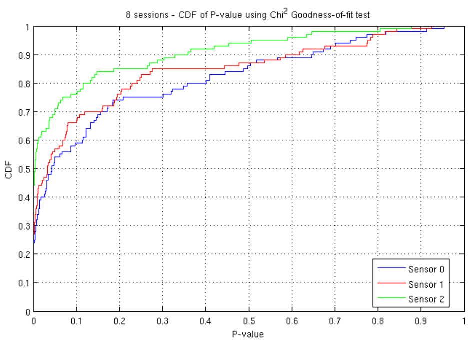


(a) 8 sessions

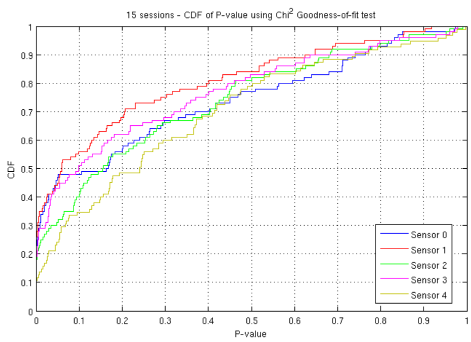


(b) 15 sessions

Figure 6.6: CDF of the MSE for different sensors

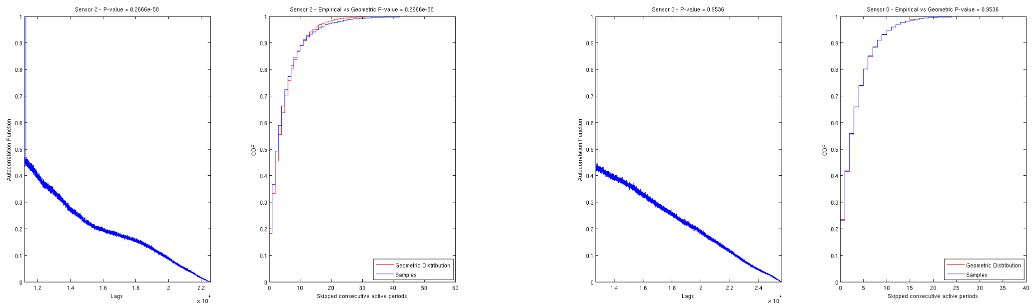


(a)



(b)

Figure 6.7: CDF of the p-value using the Chi-square Goodness-of-fit test



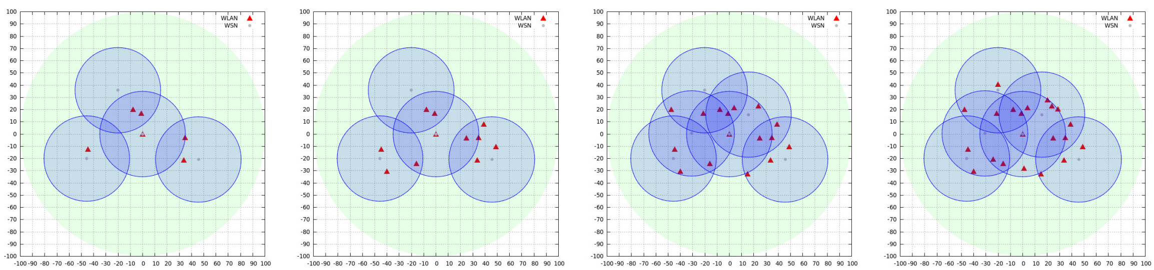
(a) (b)

Figure 6.8: Fitting of the empirical and geometric distributions for different p-values

6.3 SESSION EXPERIMENT - RESULTS

In this section we will test the estimation process carried by the Laplace Transform estimator presented in Section 2.2.2 in addition with the Kolmogorov-Smirnov validation test for different session cases.

We generated a traffic configuration that gave us a load of 30-35% for 20 users in order to be used as a fixed configuration for all the experiments. In this case, we will fix the flow number and flow inter-arrival for each session, randomizing flow sizes, packet sizes and packet inter-arrival. Using the chosen traffic configuration, we will decrease only the number of users to have four different session cases: 20, 15, 10 and 5 WLAN users. The user realizations for each session case is represented in Figure 6.9. Each one of these session-cases will represent a different load case. We run 200 runs for each session case. The traffic configuration for these experiments is presented in Table 6.1.



(a) 5 sessions - 3 sensors **(b)** 10 sessions - 3 sensors **(c)** 15 sessions - 5 sensors **(d)** 20 sessions - 5 sensors

Figure 6.9: User realization for the Local View session experiment

The experiment will be done as follows: for each run of simulation, considering a Global View model, a sensor will estimate the parameters and the Kolmogorov-Smirnov validation test for the Global View model. If the validation test is successful ($p - value > 0.05$), then each sensor, considering now the Local View model, will perform the Laplace Transform estimation with the samples gathered from the users within range and extract the parameters for the local active and idle distributions.

Table 6.2 shows the pass-rate of the K-S test for this experiment. As it can be observed, for the low-load cases, the K-S fail rate is quite high. On the other hand, this behaviour is not present in the medium and high load cases (15 and 20 session respectively).

Modeled Variable	Distribution	Parameters	
Session number	Fixed	N users arriving at simulator start	
Flow inter-arrival	Log-normal	$\mu = -1.6355$, $\sigma = 2.6286$	fixed
Flow number	Bi-Pareto	$\alpha = 0.07$, $\beta = 1.75$, $c = 295.38$, $k = 1$	fixed
Flow Size	Bi-Pareto	$\alpha = 0.00$, $\beta = 1.02$, $c = 15.56$, $k = 111$	random
Packet Size	Uniform	$min = 512$, $max = 1024$	random
Packet inter-arrival	Exponential	$\lambda = 10$	random

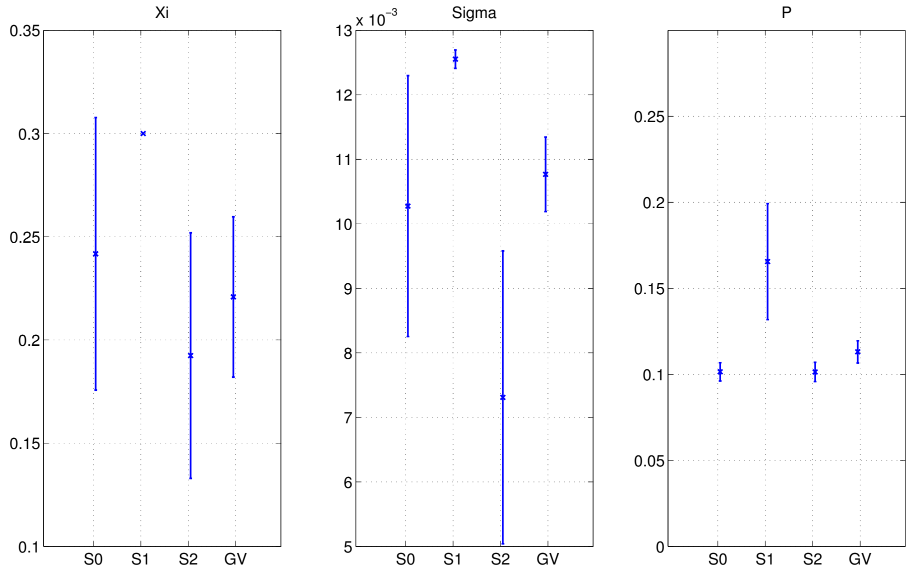
Table 6.1: Traffic configuration for the Local View session experiment according to [8].

	$P(p - value < 0.05)$
5 sessions	78.5 %
10 sessions	26 %
15 sessions	0 %
20 sessions	3,97 %

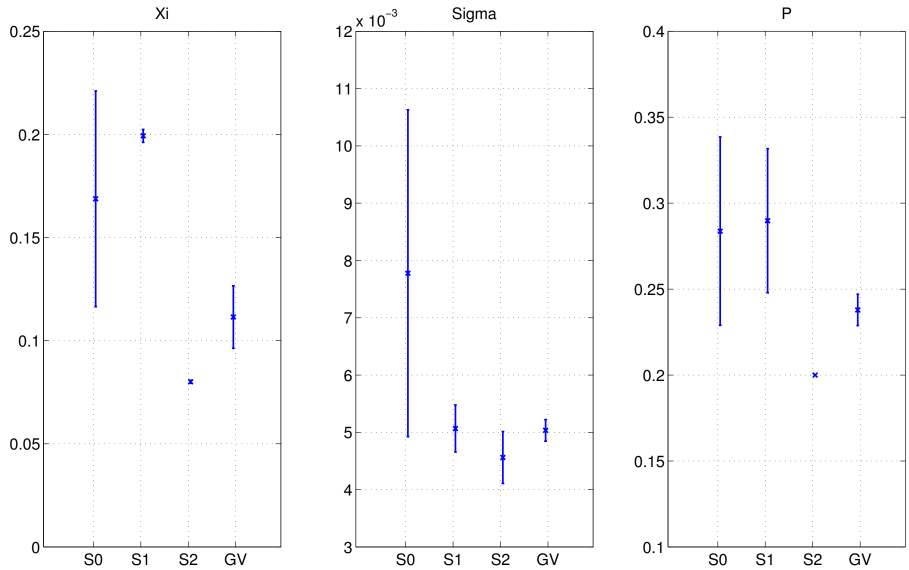
Table 6.2: K-S failure rate for the Local View session experiment (over 200 runs)

The results of the estimation for this experiment are represented in Figures 6.10 and 6.11. These figures represent only the results after the K-S filtering. Each one of the sub-figures represent each one of the estimated parameters (mean and deviation) for each one of the sensors in addition with the Global View estimated values. In these sub-plots, we represented the estimated parameters for the tests that passed previously the K-S.

It can be observed how for higher loads (see 20 and 15 session cases), the local estimated parameters fit better with the global ones than those cases with a lower load. It is possible to conclude from this experiment, that the load affects the estimation process of the local parameters. For a higher load, the sensors may be able to obtain more samples, which will lead to a closer estimation of the local parameters to the ones in the Global View model as it has been presented in the results of this section.

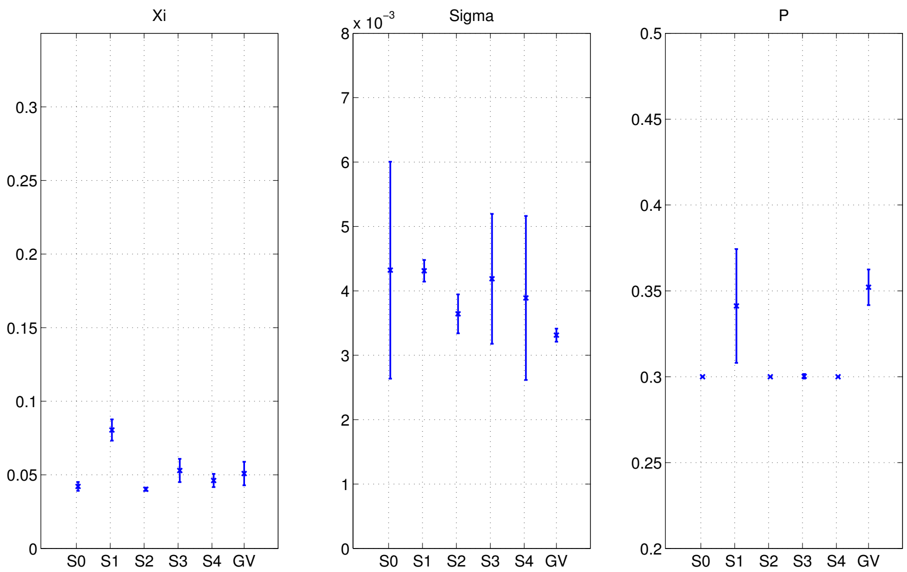


(a) 5 sessions (Load range: 8-10%)

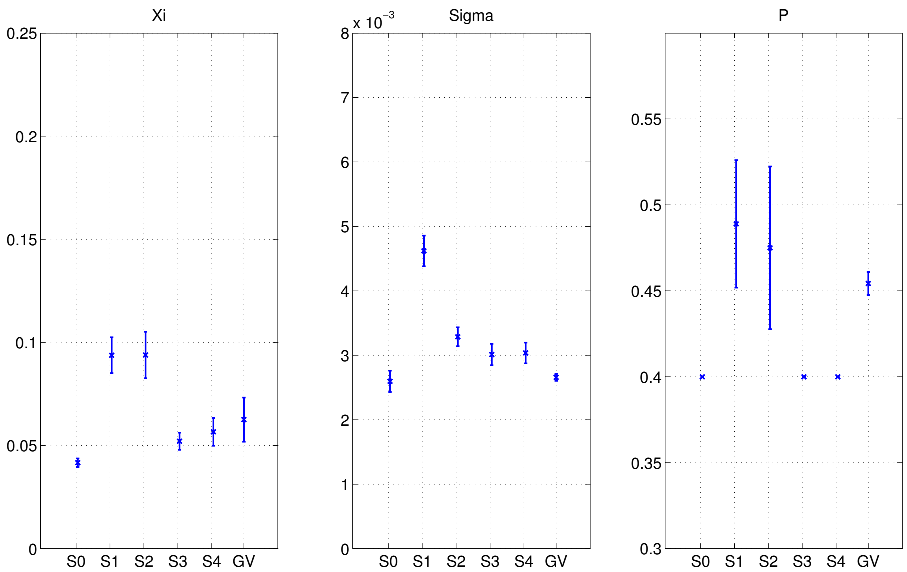


(b) 10 sessions (Load range: 12-15%)

Figure 6.10: Session experiment for the Local View model (I). Different session cases using the same traffic configuration for flow number and flow inter-arrivals.



(a) 15 sessions (Load range: 20-25%)



(b) 20 sessions (Load range: 30-35%)

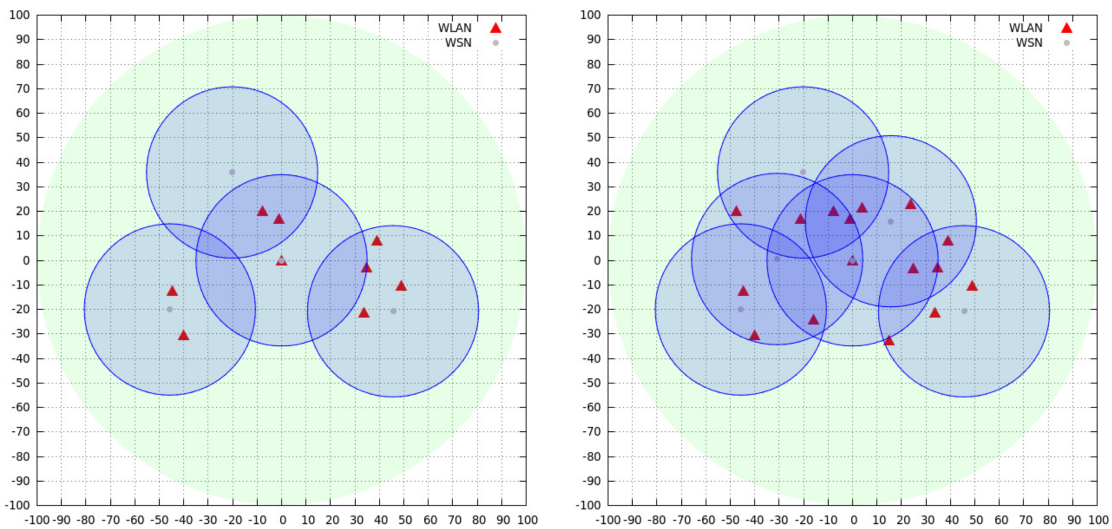
Figure 6.11: Session experiment for the Local View model (II). Different session cases using the same traffic configuration for flow number and flow inter-arrivals.

6.4 LOAD EXPERIMENT - RESULTS

In the session experiment (Section 6.4) we tested the effect of different number of sessions, representing different load cases in the estimation and validation process of the local view model. In the experiment presented in this section, we tested the effect of the load over the local view model. We studied two different load cases for two different number of sessions as follows:

- 8 sessions - ≈ 25 % load
- 15 sessions - ≈ 25 % load
- 8 sessions - ≈ 35 % load
- 15 sessions - ≈ 35 % load

The traffic configuration follows the same parametrization presented in Table 6.1. The user realization for this experiment is presented in Figure 6.12.



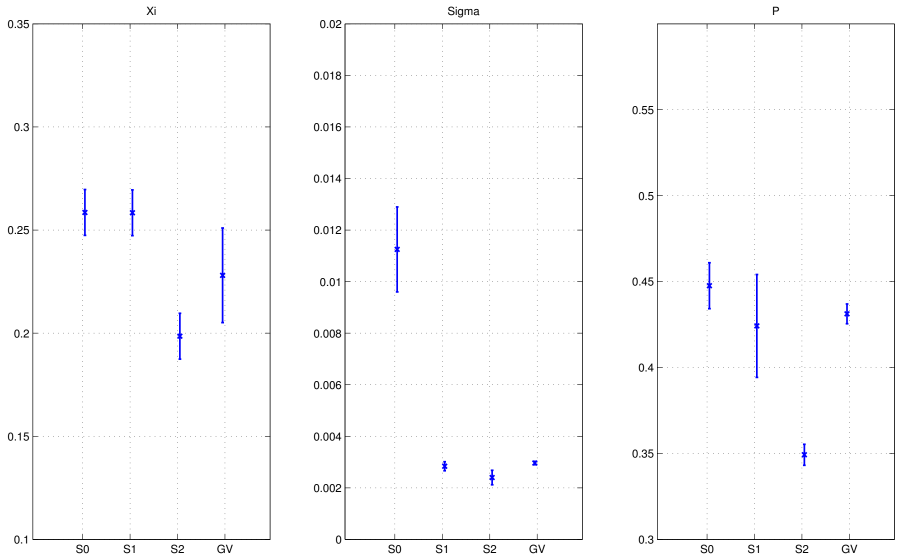
(a) 8 sessions **(b)** 15 sessions

Figure 6.12: User realization for the Local View load experiment

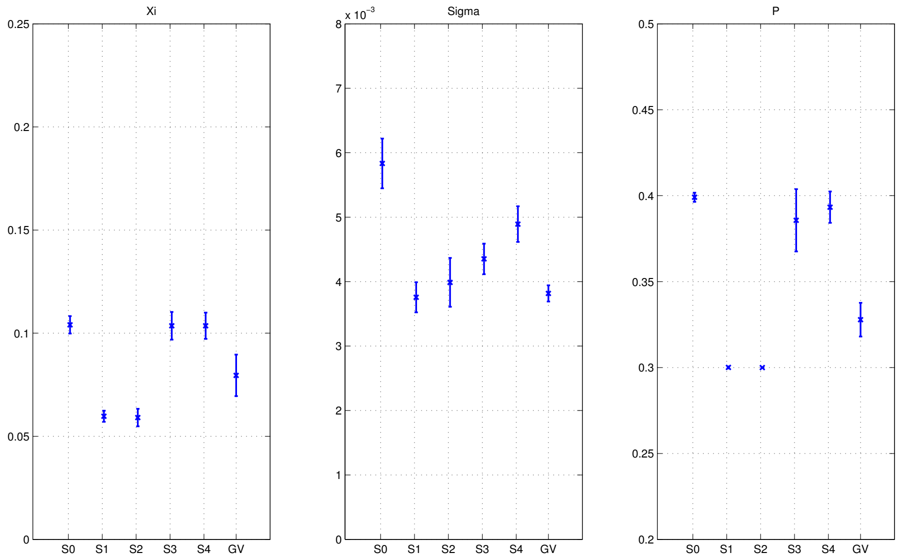
The main goal of this experiment is to compare two different load cases for two different number of sessions and observe how this affects the estimation process of the Local View model.

The results obtained for this experiment are represented in Figures 6.13 and 6.14. Figure 6.13 represents a low-medium load case for two different number of users. As it can be observed from the estimated parameters, the higher is the number of sessions, more close are the estimated local parameters to the ones of the Global View model. In Figure 6.14 we present a higher load case. Although there is still a small improvement in the 15 session case, both session cases present a similar behaviour in the estimation process when the load is higher.

From the results presented in this section, we can conclude from this experiment that the estimation process is better for a higher number of sessions when the load is low-medium. On the other hand, the estimation process is not improved significantly by the number of sessions for higher loads.

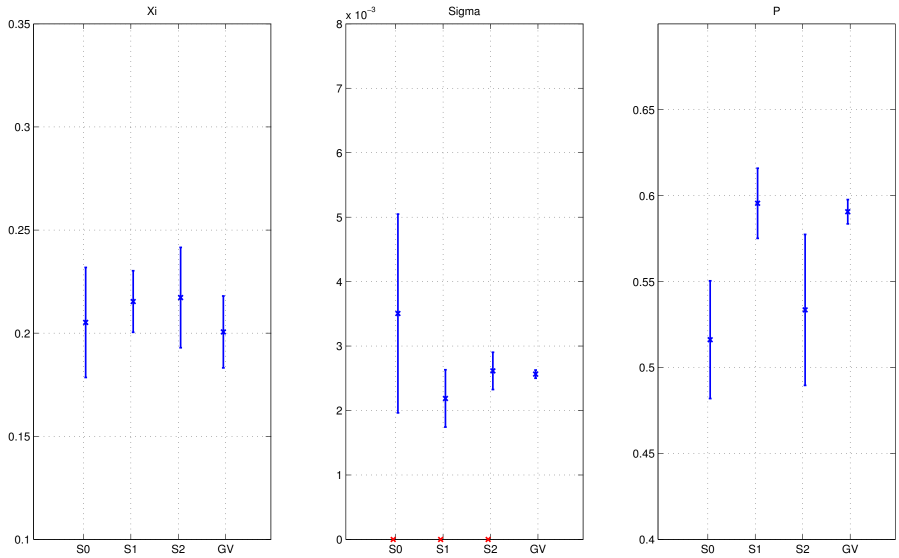


(a) 8 sessions - ≈ 25 % load

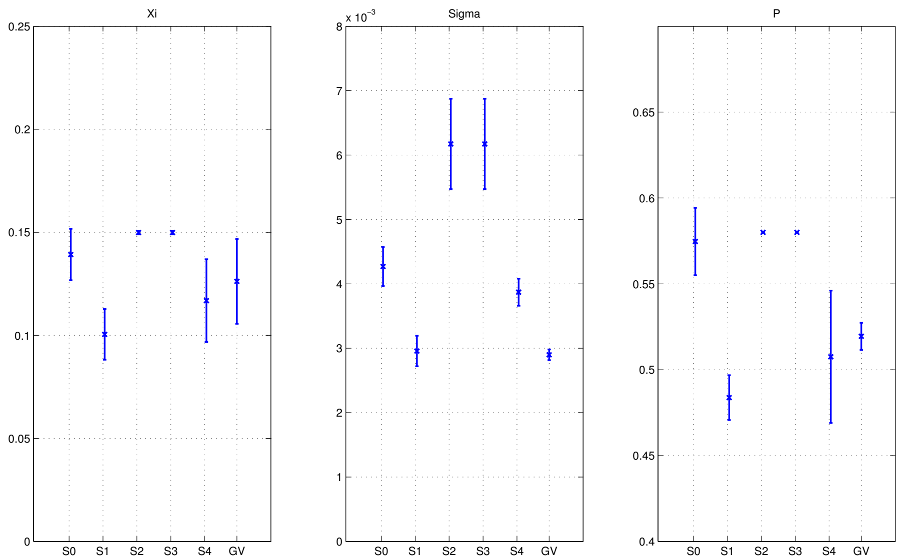


(b) 15 sessions - ≈ 25 % load

Figure 6.13: Load experiment for the Local View model (I). Different number of sessions with similar load.



(a) 8 sessions - ≈ 35 % load



(b) 15 sessions - ≈ 35 % load

Figure 6.14: Load experiment for the Local View model (II). Different number of sessions with similar load.

7 CONCLUSIONS

An extended set of tests has been developed for this project in order to determine whether and when the proposed model for the spectrum activity is suitable to be used.

As it has been explained, both idealistic and realistic approaches to model the observable load of the sensors are represented by semi-Markovian processes and the sojourn times in the Idle and Active states are defined by some determined distributions. The first step in this project was to test the validity of the proposed distributions for both states. For this, it has been proved that to model the holding times in the idle state, the proposed mixture distribution suits, showing a clear different behaviour in the distribution of the idle periods, in both idealistic and realistic models under study.

The second step was to test the estimation and validation processes in both approaches for different traffic scenarios. First of all, in the Global View model, the estimation process showed to be insensitive to the randomization of the packet level, showing that estimated parameters differed with low deviation for multiple runs of the same configuration of the higher levels (session and flow) in the multi-layer traffic model. In addition, the process was also insensitive for different number of sessions. For the validation process, the main requirement is that the active and idle distributions should be perfectly reconstructed using the estimated parameters. In order to test the quality of the reconstructed distributions and, in extension, the validity of the proposed model, a validation test such as the Kolmogorov-Smirnov test has been used. Different problems had been faced in this section of the thesis. At the beginning, the validation test chosen didn't show up to be a good tool in order to determine the quality of the reconstructed distributions. Visually, the fitting between both empirical and estimated distributions was perfect, but applying the validation test, this presented a high failure-rate. Different improvements have been implemented in the test, showing high improvement when applying the test on the truncated part of the idle distribution, avoiding applying the validation test in the uniformly distributed part of the idle distribution. Also, we presented that there is an improvement when using a high number of samples for the validation test. The results in the K-S test presented that the validation test is also insensitive for different number session cases. Finally, we wanted to determine a load region in which the proposed model can be applied with high success. The results presented that the proposed model can be applied perfectly the load region of 10 to 30 % of load, presenting a low failure rate in the validation test in addition to be insensitive to the session number.

Finally, a similar set of experiments was developed for the realistic approach (Local View). Since the mixture idle distribution followed the same behaviour in this realistic model, we proceeded with the estimation and validation tests. First of all, an analysis of the active periods was developed in order to prove the assumption in the Local View model in which is stated that the consecutive active periods are generated by totally independent sources. In addition, different validation tests were applied for this model such as Mean-Square error and Chi-squared but none of these were a good solution. In the Local View model, we implemented the estimation process using the Laplace Transform. A qualitative analysis of the local estimated parameters has been done comparing the results of the Laplace estimation process with the estimated parameters in the Global View model. The results of the experiments in this section showed that the estimation process is highly affected by the load and the number of sessions, presenting estimated parameters closer to the ones in the Global View model when the load is higher for a same number of sessions. On the other hand, a higher number of sessions also presented a small improvement in the estimated parameters.

7.1 FUTURE WORK

Multiple experiments in both approaches had been developed for this thesis. However, as the results have shown, there are still different issues to be solved. First of all it is necessary to find a proper validation tool in order to test the quality of the reconstructed distributions in both Global and Local View models, although it is more critical for the second approach. Future work may also include an extended set of experiments to determine if the proposed model can be applied in a wider region of loads in addition to the 10 to 30 % load region in which the model can be applied with high success. Finally, a deeper study of the Local View model is needed, developing tests for a more extended set of traffic scenarios in order to determine when the model is suitable to be used.

ACRONYMS

AP	Access Point
ACK	Acknowledge packet
CBR	Constant BitRate
CCA	Clear Channel Assessment
CDF	Cumulative Distribution Function
CMA	Cognitive Medium Access
CW	Contention Window
GNU	GNU's Not Unix!
GSL	GNU Scientific Library (See [2])
IEEE	Institute of Electrical and Electronics Engineers
K-S	Kolmogorov-Smirnov
MLE	Maximum Likelihood Estimation
ME	Moment Evaluation
MSE	Mean Square Error
NS	Network Simulator (See `http://isi.edu/nsnam/ns/`)
PDF	Probability Density Function
SIFS	Short Interframe Space
TCP	Transmission Control Protocol
WLAN	Wireless Local Area Netwok
WS	White Space
WSN	Wireless Sensor Network

List of Figures

List of Tables

BIBLIOGRAPHY

[1] Taylor B. Arnold and John W. Emerson. “Nonparametric Goodness-of-Fit tests for Discrete Null Distributions”. In: *The R Journal* (2011) (cit. on p. 50).

[2] M. Galassi et al. *Gnu Scientific Library: Reference Manual*. 3rd. Network Theory Ltd., 2009. ISBN: 0954612078, 9780954612078. URL: `http://www.gnu.org/software/gsl/` (cit. on p. 61).

[3] S. Geirhofer and L. Tong. “A measurement-based model for dynamic spectrum access”. In: *Proceedings of the IEEE Conference on Military Communications*. MILCOM 2006. Washington, DC, USA, Oct. 2006, pp. 1–7. DOI: `10.1109/MILCOM.2006.302405` (cit. on pp. 1, 3, 10).

[4] S. Geirhofer, L. Tong, and B. M. Sadler. “Dynamic spectrum access in WLAN channels: empirical model and its stochastic analysis”. In: *Proceedings of the first international workshop on Technology and policy for accessing spectrum*. TAPAS ’06. Boston, Massachusetts: ACM, 2006. ISBN: 1-59593-510-X. DOI: `http://doi.acm.org/10.1145/1234388.1234402`. URL: `http://doi.acm.org/10.1145/1234388.1234402` (cit. on pp. 1–4, 9, 10, 30).

[5] S. Geirhofer, L. Tong, and B.M. Sadler. “Cognitive Medium Access: Constraining Interference Based on Experimental Models”. In: *IEEE Journal on Selected Areas in Communications* 26.1 (Jan. 2008), pp. 95–105. ISSN: 0733-8716. DOI: `10.1109/JSAC.2008.080109` (cit. on pp. 1, 3).

[6] I. Glaropoulos. “Parameter Estimation of Partially Observed WLAN Spectrum Activity via Discrete Stochastic Optimization”. In: Stockholm, Sweden, Oct. 2011 (cit. on pp. 14, 15).

[7] I. Glaropoulos, V. Fodor, L. Pescosolido, and C. Petrioli. “Cognitive WSN Transmission Control for Energy Efficiency under WLAN Coexistence”. In: *Proceedings of 6th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*. CrownCom 2011. Osaka, Japan, June 2011 (cit. on pp. 2, 4, 10, 27, 28).

[8] F. Hernández-Campos, M. Karaliopoulos, M. Papadopouli, and H. Shen. “Spatio-temporal modeling of traffic workload in a campus WLAN”. In: *Proceedings of the 2nd annual international workshop on Wireless internet*. WICON ’06. Boston, Massachusetts: ACM, 2006. ISBN: 1-59593-510-X. DOI: `http://doi.acm.org/10.1145/1234161.1234162`. URL: `http://doi.acm.org/10.1145/1234161.1234162` (cit. on pp. 1–4, 17, 18, 23, 27, 31, 34, 39, 45, 53).

[9] M. Lagana. “Cognitive WSN Access Based on Local WLAN Traffic Estimation”. In: Stockholm, Sweden, 2011 (cit. on pp. 5, 9, 12–14, 17, 19, 23–25).

[10] M. Lagana, I. Glaropoulos, V. Fodor, and C. Petrioli. “Modeling and Estimation of Partially Observed WLAN Activity for Cognitive WSNs”. In: *Proceedings IEEE Wireless Communications and Networking Conference*. Accepted. Apr. 2012 (cit. on pp. 1–4, 10, 12–14, 27, 28, 34, 45).

[11] *NS-Miracle: Multi-InteRfAce Cross-Layer Extension library for the Network Simulator*. Dipartimento di Ingegneria dell’Informazione - Università degli Studi di Padova. URL: `http://telecom.dei.unipd.it/pages/read/58/` (cit. on p. 23).